Rational design of sequence-specific DNA ligands for artificial control of gene expression

Claude HELENE

Laboratoire de Biophysique, Muséum National d'Histoire Naturelle, INSERM U.201, CNRS URA 481, 43 rue Cuvier, 75231 Paris Cedex 05, France

Abstract. Oligonucleotides can be designed to bind to oligopurine.oligopyrimidine sequences of double-helical DNA. Binding occurs in the major groove and a triple helix is formed. Bifunctional oligonucleotide-intercalator conjugates have an increased affinity due to intercalation either at the triplex-duplex junction or within the triple-helical structure. Some intercalating agents can be chemically or photochemically activated to induce irreversible reactions at the target sequence. Oligonucleotide directed triple-helix formation can be used to control gene transcription. Chimeric oligonucleotides can also be designed to form a triple helix on a single-stranded target. These "oligonucleotide clamps" can block nucleic acid processing enzymes.

INTRODUCTION

The rational design of sequence-specific ligands of nucleic acid is an active field of research with several goals i) to provide molecular biologists with new tools to investigate the function of specific genes and the role of specific sequences in the control of gene expression, ii) to develop sequence-specific artificial nucleases that could cleave long DNA fragments at selected sites for, e.g., mapping genes on chromosomes, iii) to provide a rational basis for the development of new therapeutic agents based on selective modulation of gene expression.

Oligonucleotides and analogues have received considerable attention during the past few years, as a very versatile class of reagents for sequence-specific recognition and modification of nucleic acids. In the so-called "antisense" strategy an oligonucleotide is targeted to a specific messenger RNA (mRNA). It recognizes its target by hydrogen bonding interactions, forming Watson-Crick base pairs with its complementary sequence. Translation of the mRNA is inhibited according to several mechanisms (see reference 1 for a review). The oligonucleotide should have a minimum length to recognize a single mRNA species within a living cell. Depending on the base sequence a length between 12 and 15 nucleotides should be sufficient to achieve specificity in human cells.

In the "antigene" strategy the oligonucleotide is targeted to double-helical DNA (see reference 2 for a review). It recognizes Watson-Crick base pairs by hydrogen bonding interactions within the major groove and forms a local triple-helical structure. Triple helix formation may inhibit transcription of the targeted gene. Covalent modifications have been introduced into triple helix-forming oligonucleotides in order i) to stabilize triple-helical structures, ii) to induce irreversible modifications of the target sequence, iii) to create artificial sequence-specific nucleases. The basic principles and recent developments in this area of research are presented below. The reader is referred to a recent review (3) for a more detailed description of some of these aspects.

Oligonucleotide-directed sequence-specific recognition of double-helical DNA

The formation of triple helices was first discovered with homopolyribonucleotides (4). But it is only recently (5,6) that the recognition of specific DNA sequences by triple-helix-forming oligonucleotides was first described. Nucleic acid bases can form specific hydrogen bonds with purine bases already engaged in Watson-Crick hydrogen bonding interactions with complementary bases (Figure 1). These non-Watson-Crick hydrogen bonds were first observed by Hoogsteen in co-crystals of adenine and thymine derivatives (7). The interactions presented in Figure 1 are often referred to as Hoogsteen hydrogen bonds for the left column and reverse Hoogsteen hydrogen bonds for the right column. These hydrogen bonding interactions allow for recognition of T.A and C.G base-pairs from the major groove of double-helical DNA. Oligonucleotides containing thymines and cytosines, thymines and guanines, adenines and guanines can form a triple helix with DNA, provided all purines of the recognition sequence are on the same strand of DNA. The orientation of the third strand depends on the base sequence. If all nucleotides are in the anti conformation the (C,T)-containing oligonucleotide is oriented parallel to the oligopurine target sequence. A (G,A)-containing oligonucleotide has an antiparallel orientation (8), whereas the orientation of a (G,T)-containing oligonucleotide depends on the number of GpT and TpG steps (9). This arises because C.GxG and T.AxT base triplets are not isomorphous neither in the Hoogsteen nor in the reverse Hoogsteen configuration (Figure 2). This lack of isomorphism induces a distortion of the third strand backbone when moving from a T.AxT to a C.GxG base triplet along the triple helix. Hoogsteen base triplets (parallel orientation) are energetically more favorable but lead to a larger distortion of the third strand backbone. The energy penalty associated with this distortion is such that an oligonucleotide with a large number of GpT and TpG steps prefers to bind in an antiparallel orientation with reverse Hoogsteen hydrogen bonding interactions (9).

As shown on Figure 1, cytosine must be protonated to form two hydrogen bonds with a C.G base pair. This requirement induces a destabilization of triple helices containing C.GxC$^+$ base triplets when the pH is raised. Several modified bases that do not require protonation have been proposed to replace cytosines (see reference 3 for review).

Oligonucleotide-intercalator conjugates

Stabilization of triple-helical complexes
Triple-helical complexes formed by oligonucleotides with double-helical DNA are less stable than the double-helical complexes formed by the same oligonucleotides bound to complementary single-stranded sequences. In order to increase stability we have covalently attached intercalating agents at one (or both) end(s) of triplex-forming oligonucleotides. Intercalation takes place at the junction between the triplex and duplex structures and strongly stabilizes triple-helical complexes (10) (Figure 3). We have recently discovered polycyclic aromatic molecules such as benzol(e)pyrido indole derivatives which intercalate more strongly in triple helices than in double helices (11). When covalently attached to a triplex-forming oligonucleotide these intercalating agents strongly stabilize the triple-helical complex. An oligonucleotide-intercalator conjugate has been shown to inhibit transcription both *in vitro* and in cell culture (12) when targeted to the promoter region of the α subunit of the interleukin 2 receptor. Triple helix formation prevents DNA recognition by an essential protein (NF-κB) the binding of which is required to activate transcription from this promoter.

Site-specific cleavage of target DNA
When 1,10-phenanthroline was covalently attached to the 5'-end of a triplex-forming oligonucleotide, cleavage of both strands of the target DNA was observed at the triplex-duplex junction in the presence of Cu^{2+} ions and a reducing agent (13). The asymmetric distribution of
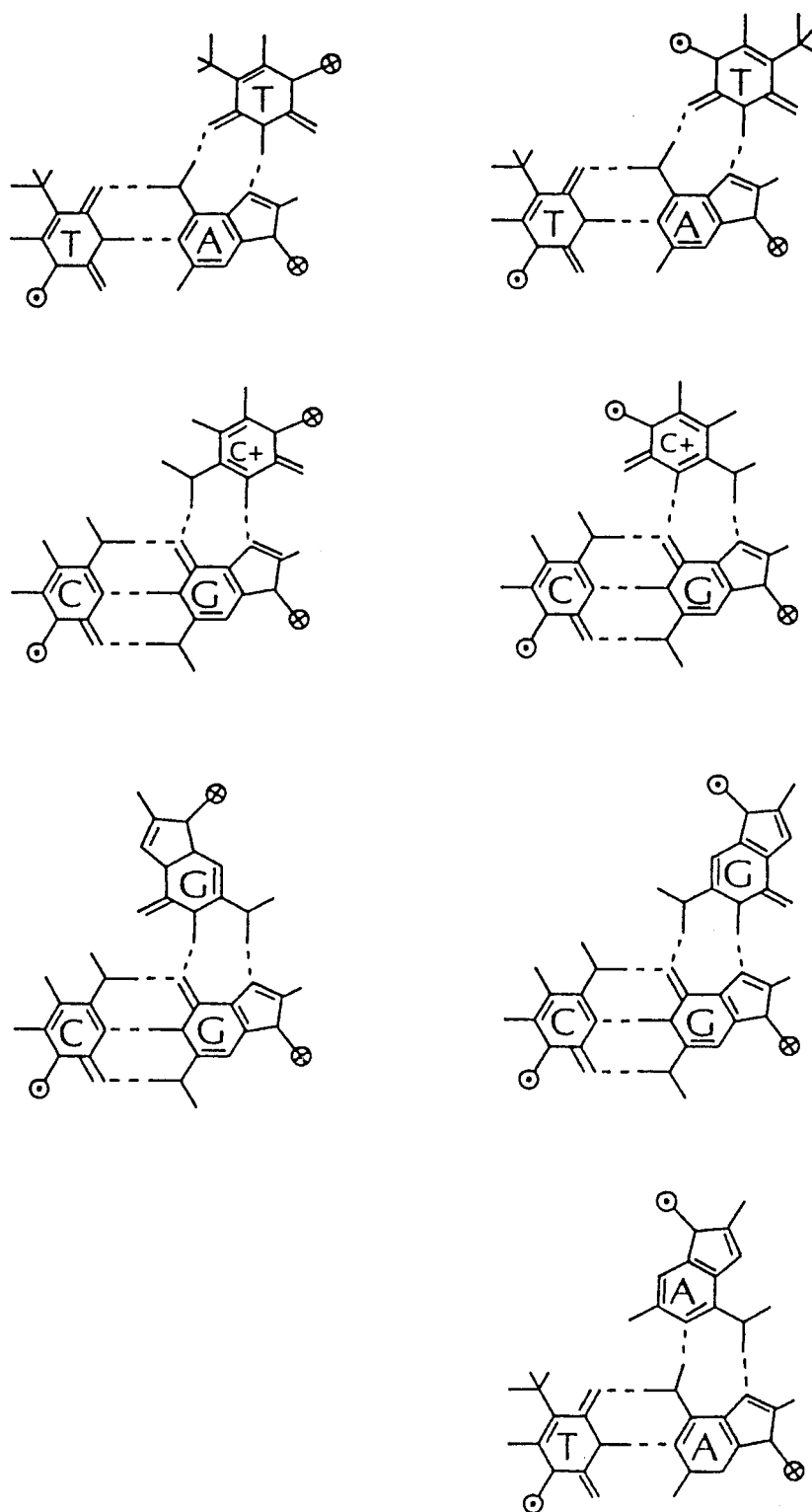
Figure 1 : Base triplets that can be formed by nucleic acid bases with Watson-Crick T.A and C.G base pairs. The left column corresponds to Hoogsteen hydrogen bonding interactions and the right column to reverse Hoogsteen interactions. Groups on the third strand involved in hydrogen bonding are NH$_2$ (C$^+$,G,A), NH(T,C$^+$,G), C = O (T) and N(A) (from reference 33).
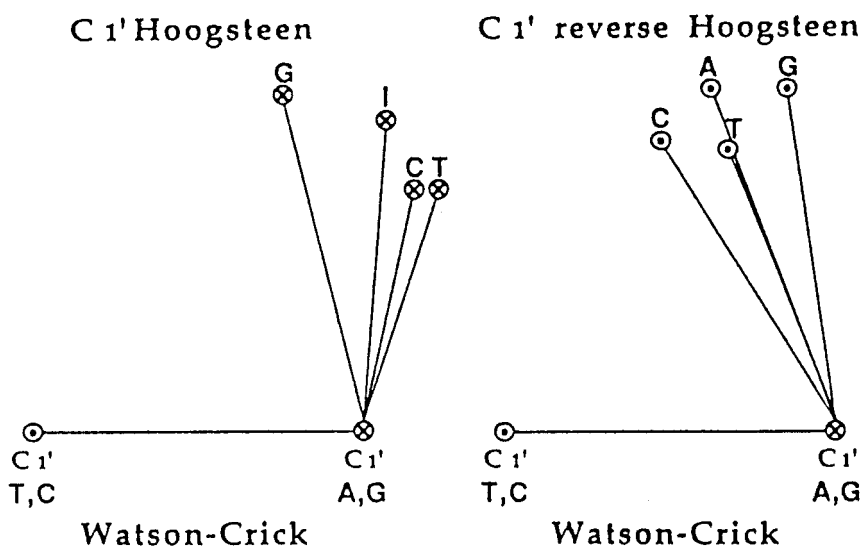
C 1' Hoogsteen   C 1' reverse Hoogsteen

Figure 2 : Position of the C1' atoms of nucleotides in the third strand with respect to Watson-Crick T.A and C.G base pairs. Only T.AxT and C.GxC$^+$ base triplets in the Hoogsteen configuration are isomorphous.

Stabilizing
intercalating agent

Cleaving reagent

Photo-induced
cleavage

Photo-induced
crosslink

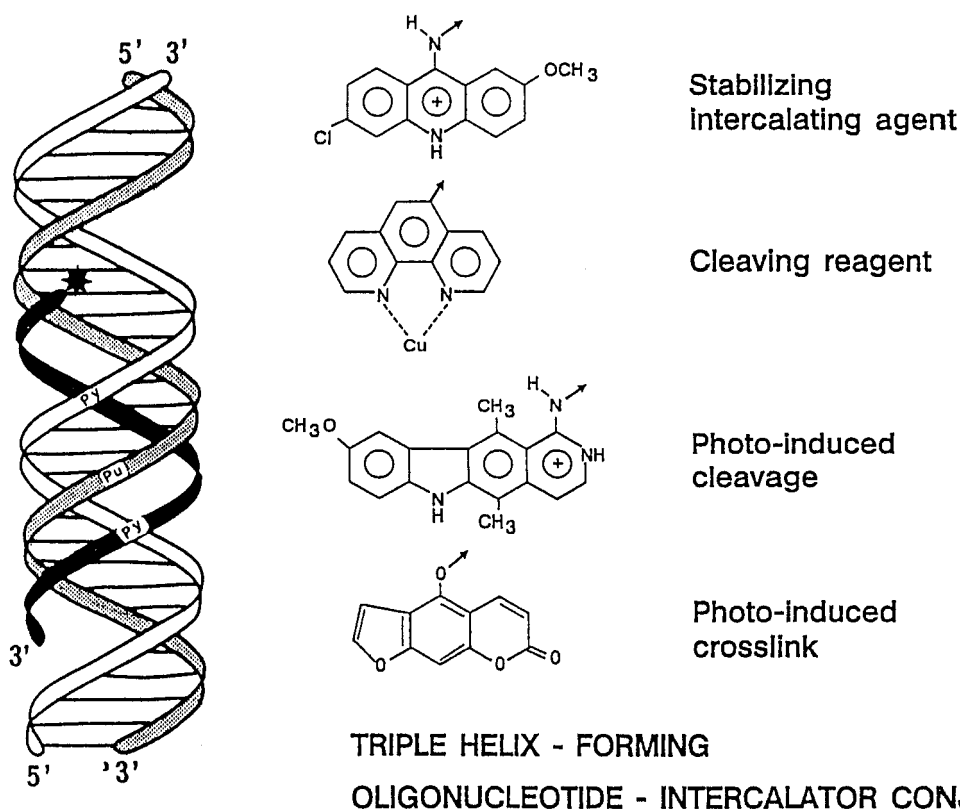TRIPLE HELIX - FORMING

OLIGONUCLEOTIDE - INTERCALATOR CONJUGATES

Figure 3 : A triple helix formed by a pyrimidine oligonucleotide (black ribbon) covalently linked to an intercalating agent (star). The intercalating agent binds at the triplex-duplex junction and can i) stabilize the triple-helical structure (acridine derivative, ref. 10), ii) induce double strand cleavage (phenanthroline, ref. 13), iii) photo-induce cleavage (ellipticine derivative, ref. 22), iv) photo-induce cross-linking of the two DNA strands (psoralen, ref. 17).

the cleavage sites on the two strands of the double helix indicated that cleavage occurred from the minor groove even though phenanthroline was brought into the major groove by the third strand oligonucleotide. A study of different linker lengths led to the conclusion that phenanthroline was intercalated at the triplex-duplex junction in such a way that the two nitrogen atoms (position 1 and 10) were pointing into the minor groove where copper chelation and oxidative reactions took place. The efficiency of double strand cleavage was ~ 70 % and the reaction was highly selective for the targeted double-helical site since no other cleavage site was detected on DNA fragments as long as 11,000 base pairs. The high selectivity of triple helix formation was also clearly demonstrated by Strobel *et al.* (14-16) using Fe-EDTA as a cleaving reagent attached to a triplex-forming oligonucleotide or after protection of the triple-helix site against methylation by a sequence-specific methylase followed by restriction enzyme cleavage. They could show that a yeast chromosome and a large fragment of a human chromosome could be cleaved at a single site. Therefore oligonucleotide-directed triple helix formation has the potential of targeting a single site within megabase size DNA.

Site-specific cross-linking of DNA strands

A psoralen-oligonucleotide conjugate was designed in such a way that psoralen could intercalate at the triplex-duplex junction in the appropriate orientation to form covalent bonds (cyclobutane rings) with two thymines on opposite strands, provided a TpA sequence was present at the junction. The reaction was triggered by UV irradiation and proved to be highly selective of the targeted site on DNA. The yield of this photoinduced cross-linking reaction was quite high (> 80 %) (17). When targeted to the promoter of the interleukin 2 receptor gene an oligonucleotide-psoralen conjugate was able to block transcription factor binding and inhibited transcription from this promoter (18).

TRIPLE HELIX FORMATION ON A SINGLE-STRANDED TARGET

An oligopurine sequence on a single-stranded nucleic acid can be used as a target for both a Watson-Crick oligopyrimidine complementary sequence and a Hoogsteen oligopyrimidine which binds as a third strand. The two oligopyrimidine sequences can be linked together and the resulting OLO (Oligo-Loop-Oligo) binds much more tightly to the oligopurine target sequence than the two separated oligomers (19). Circular oligonucleotides bind even more tightly to an oligopurine sequence (20). However the OLO can be engineered to create a duplex-triplex junction (by lengthening the Watson-Crick part) and covalently attached to an intercalating agent at its 5'-end. Intercalation takes place at the duplex-triplex junction and strongly stabilizes the complex (Figure 4). These "oligonucleotide clamps" offer several possibilities. If a psoralen is used as an intercalating agent it is possible to photo-induce covalent attachment of all three strands at a predetermined site with high efficiency. Even in the absence of covalent linkage the triple helical complex stabilized by the intercalating agent can act as a "roadblock" for nucleic acid processing enzymes such as DNA polymerases (21).

PERSPECTIVES

Oligonucleotides and their analogues represent a very versatile class of sequence-specific DNA ligands. Only purines of purine.pyrimidine Watson-Crick base pairs are recognized when natural bases are used. The formation of two hydrogen bonds per base pair provides a high selectivity of binding. It has been calculated (1) that, on a statistical basis, an oligonucleotide of about 17 bases should be able to recognize a single site within the human genome. This corresponds to the formation of 34 hydrogen bonds. However other hydrogen bonding sites are still available within the major groove and could be used if base analogues were introduced in the third strand. Alternatively one might combine an oligonucleotide with another ligand binding either the minor groove or the new groove that lies between the third strand and the
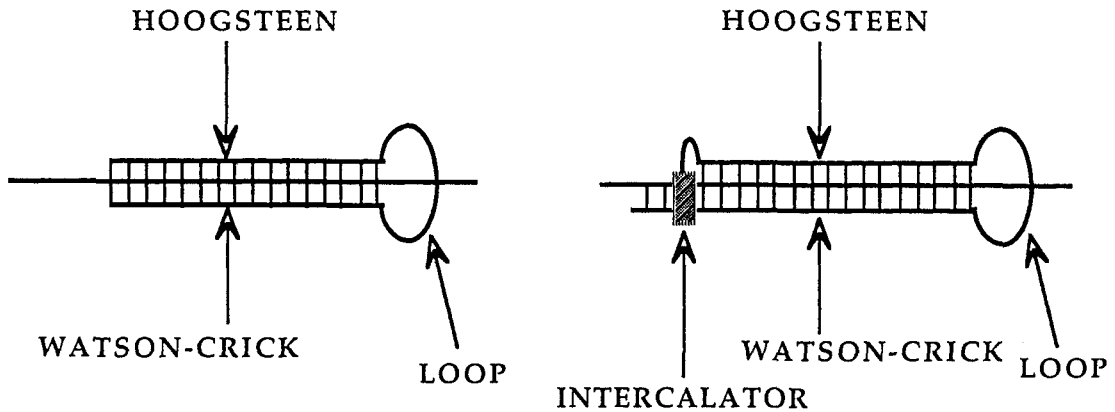
OLIGONUCLEOTIDE   CLAMPS

Figure 4 : Oligonucleotide clamps. An oligonucleotide can bind to an oligopurine single-stranded target via both Watson-Crick <u>and</u> Hoogsteen hydrogen bonds (left). An intercalating agent attached to the 5'-end can intercalate at the triplex-duplex junction provided the Watson-Crick part is lengthened by one or two nucleotides (right).

pyrimidine-rich Watson-Crick strand. Another area of interest involves derivatization of the oligonucleotide third strand to confer new properties to the triple-helical complex such as chemical modification or cleavage of the target double-helical DNA. Some possibilities have been described above. But there is a need for reagents that could induce hydrolytic rather than oxidative cleavage of DNA. Also the development of photoactivatable cleaving reagents would be of interest for some applications. An ellipticine-oligonucleotide conjugate was previously shown to photo-induce cleavage of double-helical DNA but the yield remains low (22).

The oligonucleotide backbone may also be modified in order to confer nuclease resistance for *in vivo* applications and to enhance the stability of triple-helical complexes. The recent demonstration that oligonucleotides containing 2'-O-methyl derivatives of C and U bind more strongly to double-helical DNA than 2'-deoxy analogues opens some interesting new developments (23,24). The study of RNA-containing triple helices is also of importance for the recognition of RNA.DNA and RNA.RNA double-helical structures (25-27). The survey presented above was not intended to cover all aspects of oligonucleotide-directed triple helix formation. The kinetics of triple helix formation has been shown to be much slower than that of double helix formation (28,29). Changes in association/dissociation kinetics must be investigated when new backbones and/or new bases are introduced in the third strand. The above discussion was limited to target sequences where all purines belong to the same strand of the double helix. There have been several attempts to extend the range of recognition sequences for triple helix formation. A nucleotide analogue has been recently described which allows for recognition of oligopurine sequences which are interrupted by a pyrimidine (30). It is possible to design oligonucleotides that recognize oligopurine sequences that alternate on the two strands of DNA (6,31,32). However it still remains a challenge to chemists to design base analogues that will recognize all four base pairs (A.T, T.A, C.G, G.C) when one moves along the major groove of the DNA double helix. This research area is of primary importance not only to provide new and selective tools for molecular and cellular biology but also to provide a rational basis for the development of a new therapeutic approach aimed at selectively controlling the expression of specific genes (at the transcriptional level) rather than the activity of the products of their expression.

REFERENCES

1. C. Hélène and J.J. Toulmé, <u>Biochimica et Biophysica Acta</u>, **1049**, 99-125 (1990).
2. C. Hélène, <u>Anti-Cancer Drug Design</u>, **6**, 569-584 (1991).
3. N.T. Thuong and C. Hélène, <u>Angewandte Chemie Int. Ed. Engl.</u>, **32**, 666-690 (1993).
4. G. Felsenfeld, D. Davies and A. Rich, <u>J. Am. Chem. Soc.</u>, **79**, 2023 (1957).
5. T. Le Doan, L. Perrouault, D. Praseuth, N. Habhoub, J.L. Decout, N.T. Thuong, J. Lhomme and C. Hélène, <u>Nucleic Acids Res.</u>, **15**, 7749-7760 (1987).
6. H.E. Moser, P.B. Dervan, <u>Science</u>, **238**, 645-650 (1987).
7. K. Hoogsteen, <u>Acta Crystallogr.</u>, **12**, 822-823 (1959).
8. P.A. Beal and P.B. Dervan, <u>Science</u>, **251**, 1360-1363 (1991).
9. J.S. Sun, T. de Bizemont, G. Duval-Valentin, T. Montenay-Garestier and C. Hélène, <u>C. R. Acad. Sci. Paris</u>, **313, Série III**, 585-590 (1991).
10. J.S. Sun, J.C. François, T. Montenay-Garestier, T. Saison-Behmoaras, V. Roig, N.T. Thuong and C. Hélène, <u>Proc. Natl. Acad. Sci. USA</u>, **86**, 9198-9202 (1989).
11. J.L. Mergny, G. Duval-Valentin, C.H. Nguyen, L. Perrouault, B. Faucon, M. Rougée, T. Montenay-Garestier, E. Bisagni and C. Hélène, <u>Science</u>, **256**, 1681-1684 (1992).
12. M. Grigoriev, D. Praseuth, P. Robin, A. Hemar, T. Saison-Behmoaras, A. Dautry-Varsat, N.T. Thuong, C. Hélène and A. Harel-Bellan, <u>J. Biol. Chem.</u>, **267**, 3389-3395 (1992).
13. J.C. François, T. Saison-Behmoaras, C. Barbier, M. Chassignol, N.T. Thuong and C. Hélène, <u>Proc. Natl. Acad. Sci. USA</u>, **86**, 9702-9706 (1989).
14. S.A. Strobel and P.B. Dervan, <u>Nature</u>, **350**, 172-174 (1991).
15. S.A. Strobel, A. Doucette-Stamm, L. Riba, D.E. Housman and P.B. Dervan, <u>Science</u>, **254**, 1639-1642 (1991).
16. S.A. Strobel and P.B. Dervan, <u>Science</u>, **240**, 73-75 (1990).
17. a) M. Takasugi, A. Guendouz, M. Chassignol, J.L. Decout, J. Lhomme, N.T. Thuong and C. Hélène, <u>Proc. Natl. Acad. Sci. USA</u>, **88**, 5602-5606 (1991).
    b) C. Giovannangeli, N.T. Thuong and C. Hélène, <u>Nucleic Acids Res.</u>, **20**, 4275-4281 (1992).
18. M. Grigoriev, D. Praseuth, A.L. Guyesse, P. Robin, N.T. Thuong, C. Hélène and A. Harel-Bellan, <u>Proc. Natl. Acad. Sci. USA</u>, **90**, 3501-3505 (1993).
19. C. Giovannangeli, T. Montenay-Garestier, M. Rougée, M. Chassignol, N.T. Thuong and C. Hélène, <u>J. Am. Chem. Soc.</u>, **113**, 7775-7777 (1991).
20. Prakash G. and Kool E.T., <u>J. Am. Chem. Soc.</u>, **114**, 3523-3527 (1992).
21. C. Giovannangeli, N.T. Thuong and C. Hélène, <u>Proc. Natl. Acad. Sci. USA</u> in the press (1993).
22. L. Perrouault, U. Asseline, C. Rivalle, N.T. Thuong, E. Bisagni, C. Giovannangeli, T. Le Doan and C. Hélène, <u>Nature</u>, **344**, 358-360 (1990).
23. M. Shimizu, A. Konishi, Y. Shimada, H. Inoue and E. Ohtsuka, <u>FEBS Lett.</u>, **302**, 155-158 (1992).
24. C. Escudé, J.S. Sun, M. Rougée, T. Garestier and C. Hélène, <u>C. R. Acad. Sci. Paris</u>, **315, série III**, 521-525 (1992).
25. R.W. Roberts and D.M. Crothers, <u>Science</u>, **258**, 1463-1466 (1992).
26. H. Han and P.B. Dervan, <u>Proc. Natl. Acad. Sci. USA</u>, **90**, 3806-3810 (1993).
27. C. Escudé, J.S. Sun, G. Ott, M. Sprinzl, T. Garestier and C. Hélène, submitted (1993).
28. L.J. Maher, P.B. Dervan and B.J. Wold, <u>Biochemistry</u>, **29**, 8820-8826 (1990).
29. M. Rougée, B. Faucon, J.L. Mergny, F. Barcelo, C. Giovannangeli, T. Garestier and C. Hélène, <u>Biochemistry</u>, **31**, 9269-9278 (1992).
30. Kiessling L.L., Griffin L.C. and Dervan P.B., <u>Biochemistry</u>, **31**, 2829-2834 (1992).
31. B.C. Froehler, T. Terhorst, J.P. Shaw and S.N. McCurdy, <u>Biochemistry</u>, **31**, 1603-1609 (1992).
32. P.A. Beal and P.B. Dervan, <u>J. Am. Chem. Soc.</u>, **114**, 4976-4982 (1992).
33. C. Hélène, <u>Current Opinion in Biotechnology</u>, **4**, 29-36 (1993).