2. S. L. Wei and R. E. Young, *Mar. Biol.* **103**, 541 (1989).
3. K. J. Boettcher and E. G. Ruby, *J. Bacteriol.* **172**, 3701 (1990); E. A. Meighen, *Microbiol. Rev.* **55**, 123 (1991).
4. M. McFall-Ngai and M. K. Montgomery, *Biol. Bull.* **179**, 332 (1990).
5. P. A. Blake, R. E. Weaver, D. G. Hollis, *Annu. Rev. Microbiol.* **34**, 341 (1980); R. Sakazaki and A. Balows, in *The Prokaryotes*, M. P. Starr, H. Stolp, H. G. Truper, A. Balows, H. G. Schlegel, Eds. (Springer-Verlag, Berlin, 1981), vol. 2, chap. 103.
6. K.-H. Lee and E. G. Ruby, *Abstr. Annu. Meet. Am. Soc. Microbiol.* **91**, 194 (1991).
7. M. McFall-Ngai, unpublished data.
8. K. N. Nesis, *Cephalopods of the World* (T.F.H. Publications, Neptune City, NJ, 1982), p. 70.
9. Isolations of bacteria from animal light organs, or directly from seawater, were performed as described by K. H. Nealson [*Methods Enzymol.* **57**, 153 (1978)].
10. J. L. Reichelt, K. H. Nealson, J. W. Hastings, *Arch. Microbiol.* **112**, 157 (1977); S. Fukasawa and P. V. Dunlap, *Agric. Biol. Chem.* **50**, 1645 (1986).
11. J. W. Hastings, J. Makemson, P. V. Dunlap, *Symbiosis* **4**, 3 (1987).
12. S. R. Long, *Cell* **56**, 203 (1989); B. G. Rolfe and P. M. Gresshoff, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **39**, 297 (1988); L. J. Halvorson and G. Stacy, *Microbiol. Rev.* **50**, 193 (1986).
13. B. B. Finley and S. Falkow, *Microbiol. Rev.* **53**, 210 (1989); R. R. Isberg, *Science* **252**, 934 (1991).
14. K. J. Hackett, D. E. Lynn, D. L. Williamson, A. S. Ginsberg, R. F. Whitcomb, *Science* **232**, 1253 (1986); M. S. Gil-Turnes, M. E. Hay, W. Fenical, *ibid.* **246**, 116 (1989); H. A. Gordon and L. Pesti, *Bacteriol. Rev.* **35**, 390 (1971).
15. Strains of *V. fischeri* used in this study were isolated either from the light organs of sepiolid squids (four specimens of *E. scolopes* and three specimens of *E. morsei*) or a monocentrid fish (two specimens of *M. japonicus*), or directly from Hawaiian or Californian coastal seawater samples (*9*). The strains of *V. logei* and *V. vulnificus* were provided by R. Rossen. All strains of *V. harveyi* and two *P. leiognathi* and two *P. phosphoreum* strains were seawater isolates. The two other *P. leiognathi* strains and the third *P. phosphoreum* strain were isolated from the light organs of leiognathid and macrourid fishes, respectively. All of these bacterial species have growth rate optima between 25° and 30°C (*9*).
16. Twenty-four hours after exposure to a potential symbiotic bacterial inoculum, each juvenile *E. scolopes* was rinsed in sterile seawater, its incipient light organ was homogenized, and aliquots of the homogenate spread on a seawater-based agar medium to determine the extent of bacterial infection (*3*). Uninfected squids had an undetectable number of bacteria (<20) in homogenates of their light organs.
17. Juvenile squids were fixed for scanning electron microscopy in filtered seawater containing 2% paraformaldehyde and 2% glutaraldehyde, dehydrated in an ethanol series, critical-point dried, and sputter-coated with gold, before analysis on a Cambridge model 360 scanning electron microscope. For histology, animals were fixed for 24 hours in a 0.5 M sodium phosphate buffer (pH 7.5) containing 5% formaldehyde. Specimens were then dehydrated in an ethanol series, embedded in Spurr plastic resin, and sectioned at a thickness of 1.5 μm.
18. We thank M. Montgomery and R. Young for helpful discussions, R. Emlet for making possible high-speed cinematography of juvenile light organs, and J. McCosker, T. Okutani, and R. Young for help in obtaining animals for bacterial isolations. Histological sections were provided by M. Montgomery, and A. Thompson assisted with electron microscopy. Supported by NSF grant DMB-8917293 and ONR grant N00014-91-J-1357.

25 July 1991; accepted 17 October 1991

Selective Cleavage of Human DNA: RecA-Assisted Restriction Endonuclease (RARE) Cleavage

LANCE J. FERRIN AND R. DANIEL CAMERINI-OTERO*

Current methods for sequence-specific cleavage of large segments of DNA are severely limited because of the paucity of possible cleavage sites. A method is described whereby any Eco RI site can be targeted for specific cleavage. The technique is based on the ability of RecA protein from *Escherichia coli* to pair an oligonucleotide to its homologous sequence in duplex DNA and to form a three-stranded complex. This complex is protected from Eco RI methylase; after methylation and RecA protein removal, Eco RI restriction enzyme cleavage was limited to the site previously protected from methylation. When pairs of oligonucleotides are used, a specific fragment can be cleaved out of genomes. The method was tested on λ phage, *Escherichia coli*, and human DNA. Fragments exceeding 500 kilobases in length and yields exceeding 80 percent could be obtained.

MAPPING LARGE GENOMES WOULD be greatly simplified and accelerated by development of a facile method to perform sequence-specific cleavage of DNA. In previous attempts to perform sequence-specific cleavage of genomic DNA, essentially two independent strategies have been followed. The first strategy used the ability of synthetic homopyrimidine oligonucleotides to anneal to duplex homopyrimidine-homopurine tracts to form triple-helical structures and has been used by several investigators to cleave genomes as large as *Saccharomyces cerevisiae*. This approach was first used by Moser and Dervan (*1*) to cleave a plasmid by equipping the oligonucleotide with an EDTA-Fe cleavage moiety. Subsequently, other cleavage moieties were attached to homopyrimidine oligonucleotides (*2*), or guanine-rich oligonucleotides (*3*). The main disadvantage of this targeting approach is that only homopyrimidine or guanine-rich oligonucleotides have been used successfully. In a second strategy, Szybalski and co-workers cleaved *S. cerevisiae* and *E. coli* genomes at a single introduced *lac* operator site (*4*). They first methylated Hae II sites in the DNA by using the *lac* repressor to protect a Hae II site within the *lac* operator from methylation. After removal of the methylase and the repressor, the only Hae II site unmodified and available for cleavage was in the *lac* operator. The advantages of this approach were the high yield and the high specificity. The main disadvantage was that only a *lac* operator site could be cleaved, although the method has been extended to oligopyrimidine-binding sites (*5*).

In designing a general and efficient method to perform sequence-specific cleavage and targeting of genomic DNA, we have attempted to exploit the advantages of the first two strategies while avoiding the disadvantages. We have used the ability of recombinases to pair any oligonucleotide with its homologous duplex to form a three-stranded complex. This allows specific targeting of any desired sequence.

Recent work has established that recombinases such as the RecA protein can hybridize a single strand of any sequence to an intact duplex to form a novel DNA triplex in which the third strand may include both purines and pyrimidines (*6, 7*). For reasons that at this point are not clear, the protein-free structures that we have described can only be formed in good yields on the ends of duplexes (*6*). Recently, however, we have established that related (but not necessarily identical) structures coated with RecA protein can be formed and isolated in which the pairing need not be at an end of the duplex. In these structures a single strand as short as an oligonucleotide 15 bases long can be targeted to a homologous duplex (*8*). Furthermore, these complexes are resistant to

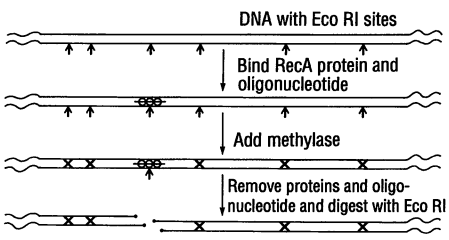

Fig. 1. Schematic of the strategy used for sequence-specific cleavage of DNA. An explanation is given in the text. This diagram shows cleavage at a single site, but a pair of oligonucleotides can be used to generate a DNA fragment. ↑, Eco RI site; **X**, methylated Eco RI site, and ⊖⊖⊖, RecA protein with an oligonucleotide.

Genetics and Biochemistry Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892.

*To whom correspondence should be addressed.

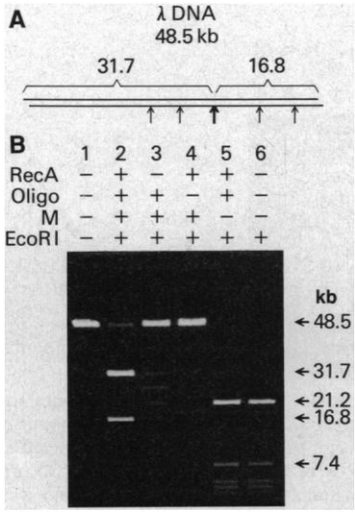

Fig. 2. Cleavage of λ DNA. (**A**) Schematic showing the position of cleavage of λ by use of an oligonucleotide homologous to the site shown by the bold arrow. Lambda DNA contains five Eco RI sites, including the one shown by the bold arrow. (**B**) Agarose gel stained with ethidium bromide showing sequence-specific cleavage of λ DNA. Lane 1 shows uncut λ DNA; lane 2 shows the complete cleavage reaction, and the other lanes show reactions with some components omitted. Unmethylated λ DNA was first protected by incubating with RecA protein and an oligonucleotide 30 bases long that was identical to the λ sequence from position 31,734 to 31,763 (*12*). After 10 min at 37°C, Eco RI methylase and *S*-adenosylmethionine were added and the reaction was allowed to proceed for 20 min. The RecA protein and methylase (M on figure) were then inactivated by heating for 15 min at 65°C. Eco RI restriction enzyme was added and the reaction was allowed to proceed for 60 min at 37°C. The reaction volume was 40 μl and contained, in order of addition: 25 mM tris-acetate, pH 7.5, 4 mM magnesium acetate, 0.4 mM dithiothreitol, 0.5 mM spermidine, 10 μg of RecA protein, 100 μM EGTA, 1.1 mM adenosine 5′-diphosphate (ADP), 0.3 mM ATP-γ-S (Fluka BioChemica), 0.18 μg of oligonucleotide, 0.9 μg of λ DNA, 4 μg of acetylated bovine serum albumin (BSA), 3.8 units of Eco RI methylase, 120 μM *S*-adenosylmethionine, and 20 units of Eco RI restriction enzyme (the last five reagents listed were from New England Biolabs). The tris-acetate, dithiothreitol, spermidine, and buffers used in the final RecA protein purification steps had been passed through Chelex 100 (Bio-Rad) columns to remove trace metal contaminants. The reactions were stopped with 5 μl of 6% sodium dodecyl sulfate (SDS), 90 mM EDTA, and 0.1% bromophenol blue. Then, 20 μl of the final reaction mixtures were mixed with 60 μl of 0.5% InCert agarose at 65°C and allowed to set in the wells of a 1.5% agarose gel. The gel was run by pulsed field electrophoresis on a CHEF-DRII system (Bio-Rad) for 36 hours at 12°C, 180 V, and 2.5-s switch time.

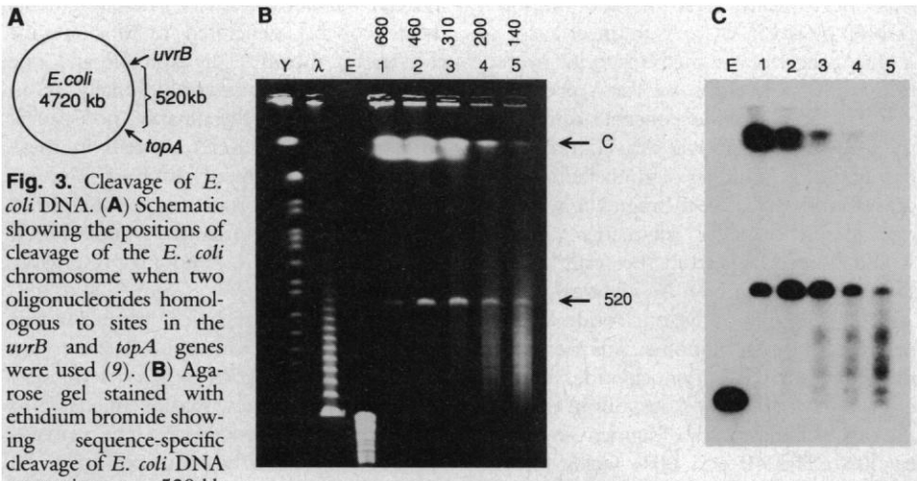

Fig. 3. Cleavage of *E. coli* DNA. (**A**) Schematic showing the positions of cleavage of the *E. coli* chromosome when two oligonucleotides homologous to sites in the *uvrB* and *topA* genes were used (*9*). (**B**) Agarose gel stained with ethidium bromide showing sequence-specific cleavage of *E. coli* DNA generating a 520-kb fragment. The compression (C) zone of the gel is also shown. Lane Y, yeast *S. cerevisiae* chromosomal DNA markers. Lane λ, λ concatamer DNA ladder. Lane E, unmodified *E. coli* DNA after a digestion to completion by Eco RI. Lanes 1 to 5, complete cleavage reactions with different amounts of oligonucleotide (in nanograms) in each lane. Reaction conditions were similar to those of Fig. 2, except the following conditions were modified to obtain good results for agarose-embedded DNA. RecA protein and oligonucleotide were preincubated with the DNA for 15 min at 37°C; methylase and S-adenosylmethionine were added and the methylation was allowed to proceed for 1 hour. The methylation was terminated by adding 100 μl of 2% SDS for 30 min at 37°C. The beads were then equilibrated in 100 mM tris-HCl, pH 8.0, 50 mM NaCl, 20 mM $MgCl_2$, 1.5 μM dithiothreitol, and nonacetylated BSA (200 μg/ml) [Calbiochem-Behring (*21*)]. Concentrations of other reagents are as in Fig. 2 except that each tube contained 20 μg of RecA protein, the indicated amount of each oligonucleotide (in nanograms), 30 μl (packed volume) of beads containing *E. coli* DNA, 40 units of methylase, and digestion was with 40 units of Eco RI restriction enzyme. After stopping the reaction, the beads were run on a 1% agarose gel for 30 hours at 12°C, 160 V, with the switch time ramped from 60 to 140 s. (**C**) Southern blot of the gel in (B). The gel was blotted onto a GeneScreen Plus nylon membrane (Dupont) according to the manufacturer's directions. We made the probe by polymerase chain reaction amplification (PCR) of a 600–base pair fragment of the *trpA* gene from *E. coli*, using ^{32}P-deoxycytidine 5′-triphosphate. The *trpA* gene lies between the *uvrB* and the *topA* gene. Lane E, which contained the same amount of DNA as lanes 1 to 5, showed the hybridization of the probe to the predicted 40-kb fragment generated by complete Eco RI digestion (*15*); the intensity of this band provided the 100% value to calculate the 520-kb fragment yield.

cleavage by several restriction endonucleases and the extent of this footprint of protection is only a few bases beyond the ends of the paired oligonucleotide.

Our specific strategy is outlined in Fig. 1. The first step in achieving sequence-specific cleavage of duplex DNA was to select a particular restriction enzyme and site for cleavage. The Eco RI system was ideal because both the restriction enzyme and methylase were free of nuclease and active under conditions optimal for RecA protein. A homologous oligonucleotide, generally 30 to 60 bases long, was synthesized such that the Eco RI site was centered in the oligonucleotide. The oligonucleotide and RecA protein (*9, 10*) were incubated with duplex DNA and the complex formed at the site of homology. Eco RI methylase and *S*-adenosylmethionine were then added and allowed to methylate all available sites, sparing the site involved in the oligonucleotide and RecA protein complex. The complex was dissociated, the methylase was inactivated, and Eco RI restriction enzyme was added to cleave at the specific Eco RI site now exposed.

Cutting at a single site is depicted in Fig. 1, but two different oligonucleotides can be added at the same time. This would allow isolation of a fragment from large or circular genomes. The following equation describes the percent yield of such a fragment:

$$(PC)^2[1 - (1 - M)C]^X(1 - N) \times 100 \quad (1)$$

where P is the efficiency of protection by RecA protein of homologous sites from methylation (a value of 1 means complete protection), C is the efficiency of restriction endonuclease cleavage, M is the efficiency of methylation of unprotected sites, X is the number of Eco RI sites contained in the fragment, and N is the fraction of fragments destroyed by nonspecific nucleases or shearing.

The three terms of the equation contain important parameters of the reaction. From the first term, it is clear that protection of the homologous site should be maximized and that drops in protection efficiency will be squared when two sites are involved. Protection efficiencies are assumed to be the same for both sites. Hundreds of trial conditions were tested to maximize protection at the desired site, while avoiding nonspecific protection at nonhomologous sites. Because of the second term, which is raised to the X power, the methylation should be carried close to completion, especially for long fragments with multiple internal Eco RI sites. Fortunately, in the buffer optimal for the most specific protection, Eco RI methylase had sufficient activity to easily obtain M values close to one. From the third term, it is clear that nonspecific nucleases

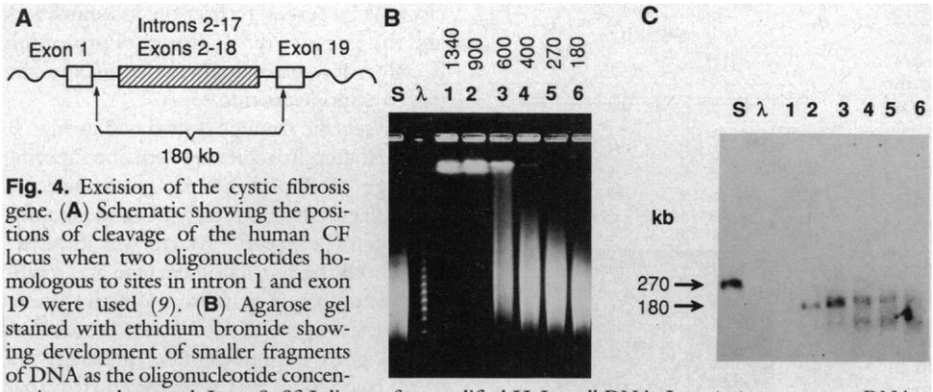

Fig. 4. Excision of the cystic fibrosis gene. (**A**) Schematic showing the positions of cleavage of the human CF locus when two oligonucleotides homologous to sites in intron 1 and exon 19 were used (*9*). (**B**) Agarose gel stained with ethidium bromide showing development of smaller fragments of DNA as the oligonucleotide concentration was decreased. Lane S, Sfi I digest of unmodified HeLa cell DNA. Lane λ, λ concatamer DNA ladder. Lanes 1 to 6, complete cleavage reactions with the indicated amount of each oligonucleotide (in nanograms). Conditions were identical to Fig. 3 except that the reaction volume at all steps was doubled, and 25 μl (packed volume) of HeLa beads were used per reaction. Sfi I (80 units; New England Biolabs) was used in lane S according to the manufacturer's directions. The beads were run on a 1% agarose gel for 32 hours at 12°C, 160 V, with the switch time ramped from 40 to 120 s. (**C**) Southern (DNA) blot of the gel in part (B). The Sfi I digest band was 270 kb long (*17*) and was used in calculating the yield of the 180-kb fragment. The probe was made by PCR of the CF cDNA T8-B3 plasmid (from the American Type Culture Collection). The probe was 550 bases long and contained 410 bases of exon 13 colinear with 140 bases of exon 14.

should be minimized, and we have found that proteins of high purity must be used. Further, large genomes, such as *E. coli* or human, must be embedded in agarose to prevent the nonspecific shearing of large fragments (*11*).

A simple demonstration of the sequence-specific cleavage of λ DNA at a single site is shown in Fig. 2. λ DNA is 48.5 kb in length and contains five Eco RI sites (*12*) (Fig. 2A). We chose to cleave at the site located at nucleotide position 31,747 to cut λ into two fragments of 31.7 and 16.8 kb. Results with an oligonucleotide 30 bases long homologous to this position are shown in Fig. 2B. Approximately 79% of the DNA was cleaved into the desired two fragments (*13*). Omission of the RecA protein resulted in incomplete methylation, possibly because of inhibition of the methylase by free oligonucleotide not coated with RecA protein. Omission of the oligonucleotide resulted in slightly incomplete methylation, presumably because of nonspecific protection from free RecA protein bound to the duplex λ DNA. This nonspecific protection showed some sequence preference, however, as some Eco RI sites were protected more than others.

To examine cleavage of *E. coli* DNA (Fig. 3), we added a pair of oligonucleotides to excise a fragment by cleavage at two sites. We chose to generate a large fragment to test the power of the method. One oligonucleotide was homologous to the *uvrB* gene, and the other to the *topA* gene (Fig. 3A). The oligonucleotides spanned Eco RI sites in each of these genes. The two genes are located 520 kb apart on the chromosome (*14*), and at least 67 Eco RI sites are between these two genes (*15*). The expected 520-kb band was observed with a yield of 40% (Fig. 3B). We observed a fairly sharp optimum at an oligonucleotide concentration of five nucleotide residues per RecA protein monomer (Fig. 3B, lane 2). This was more clearly seen in the Southern (DNA) blot (Fig. 3C). As in the cleavage of λ DNA, there was some nonspecific protection from methylation by RecA protein at lower oligonucleotide concentrations, and the 520-kb fragment was cleaved into smaller fragments. At higher oligonucleotide concentrations, the 520-kb fragment was also cleaved into smaller fragments (*16*), as would be expected from the result with λ DNA (Fig. 2B, lane 3). An identical pattern with an optimum of five nucleotide residues per RecA protein monomer was seen when the length of the oligonucleotides was increased from 30 to 60 bases, but in this case the yield of the 520-kb fragment increased to 60%. The 40 and 60% yields for the different pairs of oligonucleotides correspond to minimum single-site cutting efficiencies of 63 and 77%, respectively.

In future applications, sequence information on both sides of an Eco RI site might be difficult to obtain. We therefore measured the yield of the 520-kb fragment when the Eco RI recognition sequence, GAATTC, was at the 5′ or the 3′ end of a pair of oligonucleotides, instead of in the middle as in the experiment above. When the recognition sequence was at the 5′ end of the oligonucleotides (30 bases in length), the yield dropped two- to fourfold. When the sequence was at the 3′ end, the yield dropped an additional twofold (*16*).

Figure 4 shows the excision of a 180-kb fragment of the cystic fibrosis gene. In this gene, which has been extensively mapped and sequenced (*17–19*), one Eco RI site present in intron 1 lies 180 kb away from another Eco RI site in exon 19 (Fig. 4A). At least 41 other Eco RI sites are found within this 180-kb stretch of genomic DNA (*17*). A gel stained with ethidium bromide (Fig. 4B) shows smaller fragments generated as the concentrations of intron 1 and exon 19 oligonucleotides were reduced. This pattern was reproducible and could be used as a guide to find the optimal concentrations of oligonucleotides. The Southern (DNA) blot of the gel is shown in Fig. 4C. The greatest yield of the fragment was at 600 ng of oligonucleotide (86%). Insertion of this value into the equation given above indicates a methylation efficiency of at least 99.6%. A smaller yield (32%) was found at 900 ng, but the background cleavage was lower than at 600 ng. Thus, DNA from the 180-kb region after treatment with 900 ng probably was the most enriched in DNA from the CF locus. A predicted 48-kb fragment also was produced by specific cleavage at exons 13 and 19 (*16*). At 180 to 600 ng of oligonucleotide, the 180-kb fragment was further digested to smaller fragments. As seen in Figs. 2 and 3, some of these fragments were more prevalent than others. These fragments were probably generated by one specific cleavage at intron 1 or exon 19 and one cleavage of the fragment internally at an unmethylated site. Presumably, nonspecific binding by excess RecA protein in these samples prevented methylation of some internal sites. As the probe used is 50 kb from the exon 19 site, no fragments under 50 kb in length hybridized, although presumably they were present.

Several desirable technical features of this method deserve comment. Because the cleavage was performed by a restriction enzyme, the fragments would be amenable to standard cloning procedures. The reaction required only a few hours to perform. The two-column procedure we used yielded large amounts of RecA protein that formed more stable complexes than obtained with the commercial RecA proteins we have tried; protein preparations almost a year old have full activity. The beads were stable for at least several months. It is possible to use several of the other restriction or methylation systems available (*20*), and it might also be possible to use derivatized oligonucleotides.

This technique allows an oligonucleotide reagent to be designed to target any site in large genomes with exquisite specificity and efficiency. Such "designer" reagents have many possible uses. One application is genomic mapping. Another is the cloning and subcloning of large fragments of DNA. Ul-

timately, it may be possible to design reagents that have similar specificity and efficiency in an intracellular milieu.

REFERENCES AND NOTES

1. H. E. Moser and P. B. Dervan, *Science* **238**, 645 (1987).
2. D. Pei, D. R. Corey, P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9858 (1990).
3. P. A. Beal and P. B. Dervan, *Science* **251**, 1360 (1991).
4. M. Koob and W. Szybalski, *ibid.* **250**, 271 (1990).
5. S. A. Strobel and P. B. Dervan, *Nature* **350**, 172 (1991).
6. P. Hsieh and R. D. Camerini-Otero, *J. Biol. Chem.* **264**, 5089 (1989); P. Hsieh, C. S. Camerini-Otero, R. D. Camerini-Otero, *Genes Dev.* **4**, 1951 (1990).
7. B. J. Rao, M. Dutreix, C. M. Radding, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2984 (1991).
8. P. Hsieh, C. S. Camerini-Otero, R. D. Camerini-Otero, in preparation.
9. Oligonucleotides were purified on an FPLC Mono Q column (Pharmacia). Concentrations were determined by assuming that 1 optical density unit at 260 nm represents 33 μg. The oligonucleotide used in Fig. 2 had the sequence 5′-TCACGCCGGAAGT-**GAATTC**AAACAGGGTTC-3′ (cleavage site shown in bold). The oligonucleotides used in Fig. 3 had the sequences 5′-TCATGAGTAAACCGTTCAAACT-**GAATTC**CGCTTTTA-3′ and 5′-CGAGATCGAA-GAGGGC**GAATTC**CGCATTAA-3′. The oligonucleotides used in Fig. 4 had the sequences 5′-TAAG-TGCTCAGAAAACATTTCTTGACT**GAATTC**A-GCCAACAAAAATTTTGGGGTAGGTAG-3′ and 5′-AATGGCCAACTCTCGAAAGTTATGATTAT-TGA**GAATTC**ACACGTGAAGAAAGATGACAT-CTGG-3′.
10. We purified RecA protein using an *E. coli* strain and a detailed protocol provided by S. Kowalczykowski of the Northwestern University Medical School in Chicago (manuscript in preparation). The strain used was JC12772 [B. E. Uhlin and A. J. Clark, *J. Bacteriol.* **148**, 386 (1981)]. The purification was based on the spermidine precipitation method [J. Griffith and C. G. Shores, *Biochemistry* **24**, 158 (1985)], and used a single-stranded DNA agarose column with adenosine triphosphate (ATP) elution [M. M. Cox, K. McEntee, I. R. Lehman, *J. Biol. Chem.* **256**, 4676 (1981)] and a Mono Q column to greatly reduce trace nuclease contamination. The concentration of RecA protein was measured on the basis of an extinction coefficient of $^{1\%}E_{280} = 5.9$ [N. L. Craig and J. W. Roberts, *J. Biol. Chem.* **256**, 8039 (1981)].
11. Microbeads were used instead of agarose slabs because of the increased surface to volume ratio and greatly shortened diffusion times. Wild-type *E. coli* strain W3110 was obtained from the American Type Culture Collection and was grown overnight in Luria-Bertani medium to an OD at 600 nm of 5. Cells (5 ml) were pelleted (30 mg wet weight), washed once with 10 mM tris-HCl (pH 7.2), 20 mM NaCl, and 100 mM EDTA, and resuspended in 1 ml of this buffer. The suspension was brought to 65°C, and added to 1 ml of 1.6% low melting point agarose (InCert agarose, FMC Bioproducts) and 4 ml of paraffin oil at 65°C. Microbeads 25 to 100 μm in diameter were formed by vortexing the suspension as described [M. McClelland, *Methods Enzymol.* **155**, 22 (1987)]. Beads were digested with lysozyme and proteinase K with the ImBed kit (New England Biolabs) following the manufacturer's directions. Other lysozyme and proteinase K preparations gave equally good results. Beads were stored at 4°C and were incubated in 50 mM EDTA for 30 min and equilibrated in 25 mM tris-acetate (pH 7.5), 4 mM magnesium acetate, 0.4 mM dithiothreitol, and 0.5 mM spermidine immediately before use. Beads containing HeLa cell DNA were prepared by washing 1×10^8 cells (150 mg wet weight) twice with phosphate-buffered isotonic saline, pH 7.4, and processed as above for the *E. coli* beads, except that the lysozyme digestion step was omitted.
12. D. L. Daniels *et al.*, in *Lambda II*, R. W. Hendrix, J. W. Roberts, F. W. Stahl, R. A. Weisberg, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983), pp. 519–676.
13. Yields were calculated from densitometry of negatives of the photographs of the ethidium bromide stained gel (Fig. 2) or autoradiograms (Fig. 3), or with a Molecular Dynamics PhosphorImager (Fig. 4). Quantitation was checked at different exposures for the autoradiograms. Reproducibility from experiment to experiment was about 10%.
14. K. E. Rudd, W. Miller, J. Ostell, D. A. Benson, *Nucleic Acids Res.* **18**, 313 (1990).
15. Y. Kohara, K. Akiyama, K. Isono, *Cell* **50**, 495 (1987).
16. L. J. Ferrin and R. D. Camerini-Otero, unpublished observations.
17. J. M. Rommens *et al.*, *Science* **245**, 1059 (1989).
18. J. R. Riordan *et al.*, *ibid.* p. 1066.
19. J. Zielenski *et al.*, *Genomics* **10**, 214 (1991).
20. Sss I, Alu I, and Hha I methylases (New England Biolabs) were active under conditions optimal for the RecA protein.
21. We have confirmed the observation that this buffer is excellent for inhibiting nonspecific or star activity of Eco RI on agarose-embedded DNA [W. W. Wilson and R. M. Hoffman, *Anal. Biochem.* **191**, 370 (1990)].

22. We thank G. Aurbach, G. Felsenfeld, P. Hsieh, H. Nash, and R. Proia of NIH for reading the manuscript before submission, G. Poy for oligonucleotide syntheses, and L. Robinson for her assistance. This paper is dedicated to Gerald D. Aurbach.

29 July 1991; accepted 29 October 1991

Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide

PETER E. NIELSEN, MICHAEL EGHOLM, ROLF H. BERG, OLE BUCHARDT

A polyamide nucleic acid (PNA) was designed by detaching the deoxyribose phosphate backbone of DNA in a computer model and replacing it with an achiral polyamide backbone. On the basis of this model, oligomers consisting of thymine-linked aminoethylglycyl units were prepared. These oligomers recognize their complementary target in double-stranded DNA by strand displacement. The displacement is made possible by the extraordinarily high stability of the PNA-DNA hybrids. The results show that the backbone of DNA can be replaced by a polyamide, with the resulting oligomer retaining base-specific hybridization.

REAGENTS THAT BIND SEQUENCE specifically to double-stranded DNA are of major interest in molecular biology and could form the basis for gene-targeted drugs (*1*). Sequence-specific binding to operator DNA regions is the basis for the biological function of a large number of gene-regulatory proteins (*2*). Synthetic peptides that contain the approximately 50 amino acid residues constituting the DNA binding domain of such regulatory proteins can retain the DNA binding specificity of the parent protein (*3*), but at present it is not possible to design peptides that bind to desired DNA sequences. However, pyrimidine or purine oligonucleotides bind sequence specifically to homopurine regions of double-stranded DNA by triple helix formation through T·A-T and C^+·G-C or G·G-C and A·A-T triplets (*4*). The triple-helix principle has generally been applied to homopurine DNA targets. Furthermore, oligonucleotides are difficult to prepare in large scale (millimole to mole quantities), and introduction of modified nucleobases and conjugation to other ligands present major obstacles. One way to overcome these drawbacks would be to replace the deoxyribose phosphate backbone of DNA with a polyamide backbone that was homomorphous to DNA in terms of the number of backbone bonds and the distance between backbone and nucleobase.

We wanted to design a polyamide that could recognize double-stranded DNA through Hoogsteen-like base pairing in the major groove by nucleobases or other ligands having the proper hydrogen donor-acceptor properties. Thymine was initially chosen because it can participate in stable Hoogsteen triple helices with oligonucleotides (*4*) and because it presented the fewest synthetic obstacles. The proper distances in the backbone were estimated with a computer model by constructing a normal T·A-T triplex, removing the deoxyribose-phosphate backbone of the third (the T) strand, and building a polyamide backbone in its place. Units of 2-aminoethylglycine were found to fit when the thymine was attached through a methylenecarbonyl group (Fig. 1).

The resulting polyamide nucleic acid (designated PNA-1, Fig. 1) was equipped with a helix-threading acridine (*5*) for two pur-

P. E. Nielsen, Research Center for Medical Biotechnology, Department of Biochemistry B, The Panum Institute, Blegdamsvej 3c, DK-2200 N, Copenhagen, Denmark.
M. Egholm and O. Buchardt, Department of Organic Chemistry, The H. C. Ørsted Institute, Universitetsparken 5, DK-2100 Ø, Copenhagen, Denmark.
R. H. Berg, Polymer Group, Materials Department, Risø National Laboratory, DK-4000 Roskilde, Denmark.