Reagents for the site-specific cleavage of megabase DNA

Peter B. Dervan

The physical mapping of chromosomes will be facilitated by methods of breaking large DNA into manageable fragments, or cutting uniquely at genetic markers of interest. Key issues in the design of sequence-specific DNA cleaving reagents are the specificity of binding, the number of different sequences that can be targeted and the cleavage yield.

OLIGONUCLEOTIDE-directed triple-helix formation is one of the most versatile methods for the sequence-specific recognition of double-helical DNA[1,2]. The ability to target a broad range of DNA sequences, the high stabilities of the triple-helical complexes and the sensitivity to single base mismatches make this a powerful technique for binding single sites within large segments of double-helical DNA. This approach to DNA recognition has been used to mediate single site-specific cleavage of human chromosomal DNA[3], as well as to inhibit transcription *in vitro*[4,5].

Recognition of double-helical DNA by triple-helix formation

At least two classes of DNA triple helices exist that differ in the sequence compositions of the third strand, relative orientations and positions of the backbones of the three strands, and base triplet interactions. Pyrimidine oligonucleotides bind purine tracts in the major groove of double-helical DNA parallel to the purine Watson–Crick strand. Sequence specificity is derived from thymine (T) recognition of adenine•thymine (AT) base pairs (T•AT base triplets) and N-3 protonated cytosine (C^+) recognition of guanine•cytosine (GC) base pairs (C+GC base triplets)[6–8]. An additional family of triple-helical structures consists of purine-rich oligonucleotides bound in the major groove to purine tracts of DNA antiparallel to the Watson–Crick purine strand[4,9,10]. Sequence specificity is derived from guanine (G) recognition of GC base pairs (G•GC base triplets) and from A or T recognition of AT base pairs (A•AT and T•AT triplets)[4,9,10]. The identification of other natural base triplets[11], alternate strand triple-helix formation[12,13] and the design of non-natural base triplets[14] have increased the general application of oligonucleotide-directed triple-helix formation for the recognition of DNA sequences containing all four base pairs.

The stability of triple helices is dependent on the length, sequence composition and modification of the oligonucleotide, as well as solution conditions, including pH and cation concentrations[1–15]. A full understanding of the factors contributing to the stability and specificity of the binding of oligonucleotides to double-helical DNA will require a complete characterization of the thermodynamics of the complex formation.

Oligonucleotides conjugated to EDTA•Fe can specfically bind double-helical DNA by triple-helix formation and produce double-strand cleavage at single sites in DNA[1]. The free energy for oligodeoxynucleotide-directed triple-helix formation at a single site on a DNA plasmid fragment has been analysed using quantitative affinity cleavage titration[15]. Measurement of the amount of site-specific cleavage of a DNA restriction fragment (24 °C, 100 mM Na^+, 1 mM spermine 4HCl, pH 7.0) over a concentration change of four orders of magnitude for oligodeoxyribonucleotide-EDTA•Fe (5′-T*TTTCTCTCTCTCT-3′) yields an equilibrium binding constant, $K_T = 3.7 \times 10^6$ M^{-1} ($\Delta G_T = -9.0$ kcal mol^{-1}) for a 15-base pair purine tract. Single internal base triplet mismatches result in a destabilization of the local triple-helical structure by 2.5–3.0 kcal mol^{-1})[15].

Sequence-specific double-strand cleavage of DNA

The chemical yields for double-strand oxidative DNA cleavage by oligonucleotides equipped with EDTA•Fe are typically low (15–25%)[1]. Such yields are adequate for addressing questions related to molecular recognition, including binding specificities, site size and orientation. If the full potential of oligonucleotide-directed recognition of double-helical DNA is to be realized for site-specific cleavage of DNA of high complexity, the design of oligonucleotides equipped with moieties capable of quantitative modification of DNA is an area for further development.

Enzymatic cleavage mediated by triple-helix formation. There appears to be two separate approaches that have in common coupling enzymatic cleavage with oligonucleotide-directed recognition of DNA. One is to attach DNA cleaving enzymes to oligonucleotides[16]. An oligonucleotide-staphylococcal nuclease adduct cleaves plasmid DNA in 75% yield[16]. Although the nuclease is, in a formal sense, nondiffusible, multiple cleavage sites at the binding site are observed, likely due to conformational flexibility of the attached enzyme. The other approach is termed Achilles heel cleavage[17] and involves transient site-specific protection from enzymatic methylation mediated by triple-helix formation, followed by triple-helix disruption and cleavage by a restriction enzyme[18] (Fig. 1). Enzymatic cleavage mediated by this approach has been shown to cleave a single site in a yeast chromosome in 95% yield[18] and human chromosome 4 (200 megabase pairs in size) in 80–90% yield[3]. Triple helix-mediated enzymatic cleavage affords high specificity that can, in principle, be customized to unique genetic markers without artificial insertion of a target sequence.

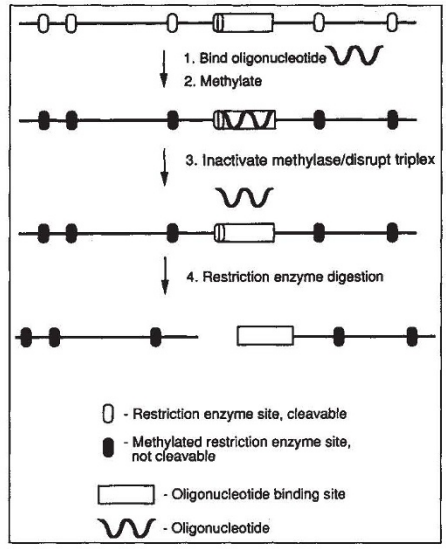

FIG. 1 General scheme for single-site enzymatic cleavage of genomic DNA by oligonucleotide-directed triple-helix formation. Chromosomal DNA is equilibrated with an oligonucleotide in a methylase-compatible buffer containing polycation. *Eco*RI methylase, which methylates the central adenines of the sequence 5′-GAATTC-3′ and renders the sequence resistant to cleavage by *Eco*RI restriction endonuclease, is added and allowed to proceed to completion. The methylase is inactivated and the triple helix is disrupted at 55 °C in a high-pH buffer containing detergent. After washing extensively, the chromosomal DNA is re-equilibrated in restriction enzyme buffer and cut to completion with *Eco*RI restriction endonuclease. The cleavage products are separated by pulsed-field gel electrophoresis and efficiencies quantitated by Southern blotting (from ref. 18).

The use of degenerate oligonucleotides in this technique to screen rapidly genetic markers for overlapping triple-helix methylation–restriction sites could make it possible to cut chromosomal DNA uniquely and efficiently at endogenous sites with minimal sequence information[3]. It remains a challenge for chemists to match the enzymatic yields for DNA cleavage with rational mechanism-based chemical approaches for quantitative modification of DNA within a triple-helical complex.

Sequence-specific alkylation of double-helical DNA by triple-helix formation. Attachment of the nondiffusible electrophile *N*-bromoacetyl to the 5 position of a thymine at the 5′ end of a pyrimidine oligodeoxyribonucleotide affords sequence-specific alkylation of a guanine two base pairs to the 5′ side of a local triple-helical complex in >96% yield[19]. *N*-bromoacetyloligodeoxyribonucleotides bind adjacent inverted purine tracts on double-helical DNA by triple-helix formation and alkylate single guanine positions on opposite strands at 37 °C (pH 7.4). After depurination, double-strand cleavage at a single site within plasmid DNA occurs in greater than 85% yield[19] (Fig. 2). The resulting DNA fragments from site-specific alkylation and cleavage can be ligated with DNA fragments generated by restriction endonuclease digestion. This nonenzymatic approach, which couples sequence-specific recognition with sequence-dependent cleavage, affords double-strand, site-specific cleavage in megabase-size DNA. A yeast chromosome was cleaved at a single site in 85–90% yield[19]. This nonenzymatic approach affords double-strand cleavage yields comparable to those attained by triple helix-mediated enzymatic cleavage of DNA. Undoubtedly, the *N*-bromoacetyloligodeoxyribonucleotide-mediated cleavage is a first-generation approach and further refinements can be expected. Because oligonucleotide-directed recognition of double-helical DNA by triple-helix formation is one million times more specific than restriction enzymes, it is not unthinkable that quantitative reactions targeted to single atoms in the human genome (3 billion base pairs) should be possible by strictly chemical methods.

Current limitations

Oligonucleotide-directed targeting of double-helical DNA by triple-helix formation is still limited to mostly purine tracts[1,2,9] or mixed sequences of the type $(purine)_m(pyrimidine)_n$ and $(pyrimidine)_m(purine)_n$[12,13] Novel bases that bind TA and CG base pairs remain to be designed or discovered to complete the recognition code. For this reason, the recently reported RecA-mediated oligonucleotide Achilles heel procedure, which appears to allow all four base-pair targeting, shows significant promise for genome research[20].

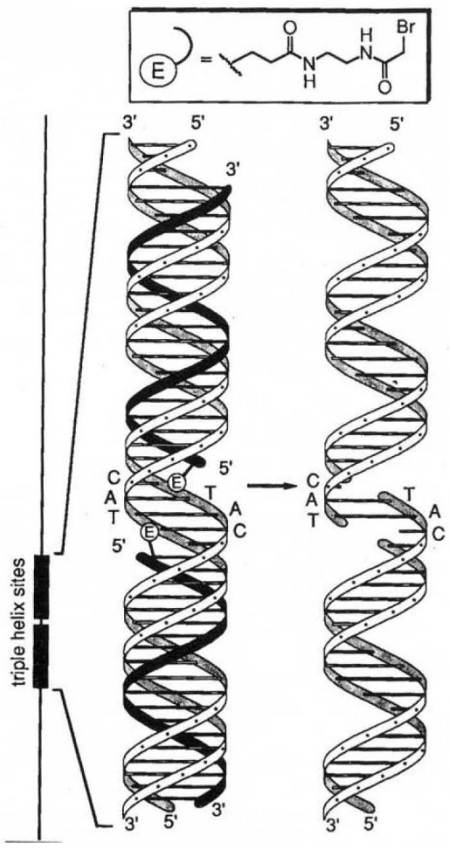

FIG. 2 Oligodeoxyribonucleotides equipped with an electrophile (E) at the 5′ end bind to adjacent inverted binding sites on double-helical DNA by triple-helix formation and alkylate at single guanine positions on opposite strands[19]. After depurination, double-strand cleavage produces sequence-specific overhangs suitable for ligation.

With regard to single-site targeting of chromosomes *in vivo*, major technical hurdles still exist that will require both chemical and biological advances, such as (1) the design and efficient synthesis of modified oligonucleotide analogues that permeate cells and are resistant to nucleases and (2) a detailed understanding of both chromatin structure and transcriptional control *in vivo*. □

Peter B. Dervan is in the Division of Chemistry and Chemical Engineering at the California Institute of Technology, Pasadena, California 91125, USA. For more information, fill in reader service number 100.

1. Moser, H. E. & Dervan, P. B. *Science* **238**, 645–650 (1987).
2. LeDoan, T. *et al. Nucleic Acids Res.* **15**, 7749–7760 (1987).
3. Stobel, S. A., Doucette-Stamm, L. A., Riba, L., Housman, D. E. & Dervan, P. B. *Science* **254**, 1639–1642 (1991).
4. Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, J. & Hogan, M. E. *Science* **241**, 456–459 (1989).
5. Maher, L. J., Dervan, P. B. & Wold, B. *Biochemistry* **31**, 70–81 (1992).
6. Felsenfeld, G., Davies, D. R. & Rich, A. *J. Am. chem. Soc.* **79**, 2023–2024 (1957).
7. Rajagopal, P. & Feigon, J. *Biochemistry* **28**, 7859–7870 (1989).
8. Radhakrishnan, I., Gao, X., de los Santos, C., Live, D. & Patel, D. J. *Biochemistry* **30**, 9022–9030 (1991).
9. Beal, P. A. & Dervan, P. B. *Science* **251**, 1360–1363 (1991).
10. Radhakrishnan, I., de los Santos, C. & Patel, D. J. *J molec. Biol.* **221**, 1403–1418 (1991).
11. Griffin, L. C. & Dervan, P. B. *Science* **245**, 967–971 (1989).
12. Horne, D. A. & Dervan, P. B. *J. Am. chem. Soc.* **112**, 2435–2437 (1990).
13. Beal, P. A. & Dervan, P. B. *J. Am. chem. Soc.* **114**, 4976 (1992).
14. Koh, J. S. & Dervan, P. B. *J. Am. chem. Soc.* **114**, 1470–1478 (1992).
15. Singleton, S. F. & Dervan, P. B. *J. Am. chem. Soc.* (in the press).
16. Pei, D., Corey, D. R. & Schultz, P. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9858–9862 (1990).
17. Koob, M., Grimes, E. & Szybalski, W. *Science* **241**, 1084–1086 (1988).
18. Strobel, S. A. & Dervan, P. B. *Nature* **350**, 172–174 (1991).
19. Povsic, T. J., Strobel, S. A. & Dervan, P. B. *J. Am chem. Soc.* **114**, 5934–5941 (1992).
20. Ferrin, L. J. & Camerini-Otero, R. D. *Science* **254**, 1494–1497 (1991).