

**Site Specific Cleavage of Genomic DNA
Mediated by Triple Helix Formation**

Thesis by
Scott A. Strobel

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1992

(Defended January 24, 1992)

Acknowledgments

This work was through the guidance and assistance of numerous people to whom I wish to express special gratitude. I would especially like to thank Dr. Peter Dervan for his enthusiasm and his vision. He presented the majority of the results described in this thesis as a potential thesis project when I arrived at Caltech. It is a tribute to his insight that many of those original ideas worked successfully. He has provided a tremendous environment to do scientific research, both by the quality of the facilities and by the quality of the people he brings to his group. Among those group members whom I would especially like to thank are Jim Maher, Tom Povsic, Peter Beal, and Karen Draths. Their helpful suggestions and example made this work possible. I also thank the whole of the Dervan group, both past and present, for their companionship and scientific excellence. I would also like to thank our collaborators at MIT, namely Lynn Doucette-Stamm, Laura Riba, and David Housman.

I wish to thank my parents for their support and example, and Herb Harvis for his generous heart and dedication to his son.

I express appreciation to the ARCS foundation and the Howard Hughes Medical Institute for supporting this work financially.

Finally, I dedicate this thesis to my family: Lynnette, Benjamin, and Sarah. They were the ones who were with me through the successes and failures of research, and helped me to keep the whole experience in its proper perspective. I have loved working on this project, but they are the true love of my life.

Abstract

Physical isolation of large segments of chromosomal DNA is a major goal of human genetics. This would be greatly assisted by a generalizable technique for the cleavage of chromosomal DNA at a single site. Pyrimidine oligonucleotide directed triple helix formation is a generalizable motif for the site specific recognition of duplex DNA. The pyrimidine oligodeoxyribonucleotide is bound in the major groove parallel to the purine strand through formation of specific Hoogsteen hydrogen bonds. Specificity is derived from thymine (T) recognition of adenine-thymine (AT) base pairs (T•AT triplet) and N3 protonated cytosine (C⁺) recognition of guanine-cytosine (GC triplet). Theoretically, the binding site size is sufficient to specify a unique site in gigabase DNA, yet the binding motif is sufficiently generalizable to recognize an endogenous site every 1000 base pairs.

This thesis describes the application of oligonucleotide directed triple helix formation to bind unique target sites in bacteriophage λ , yeast, and human genomic DNA. Cleavage at the binding sites are achieved by affinity cleaving with EDTA•Fe(II) derivatized oligonucleotides, alkylation with bromoacetyl derivatized oligonucleotides, and by site specific triple helix mediated methylase inhibition followed by digestion with the cognate endonuclease. Cleavage of genomic substrates with progressively greater complexity is described. Bacteriophage λ genomic DNA (48.5 kilobase pairs) was targeted at a single endogenous homopurine site within the origin of replication. This substrate was also used to demonstrate cooperative binding of heterologous oligonucleotides to duplex DNA at contiguous binding sites. An engineered target site on yeast chromosome III (340 kilobase pairs) was cut

quantitatively at a single site within total yeast genomic DNA (14 megabase pairs) by both chemical and enzymatic techniques.

Techniques for the identification of endogenous triple helix target sites within unsequenced genetic markers were developed and successfully used to characterize a target site on human chromosome 4, proximal to the Huntington disease gene. As a test for the site specific cleavage of gigabase DNA, this site near the end of human chromosome 4 was cleaved by triple helix mediated enzymatic cleavage. This generated a specific 3.6 Mb fragment in greater than 80% yield that contained the entire candidate region for the Huntington mutation.

Table of Contents

Acknowledgments.....	iii
Abstract	iv
Table of Contents	vi
List of Figures and Tables	viii
Chapter 1: Introduction	1
Chapter 2: Cleavage of Bacteriophage λ	34
Part I: Site Specific Cleavage of Bacteriophage Lambda	
Genomic DNA	34
Part II: Cooperative Site Specific Binding of Oligonucleotides	
to Duplex DNA	63
Chapter 3: Site Specific Cleavage of Yeast Genomic DNA Mediated	
by Triple Helix Formation	81
Part I: Site Specific Cleavage of a Yeast Chromosome by	
Oligonucleotide Directed Triple Helix Formation	83
Part II: Single Site Enzymatic Cleavage of Yeast Genomic	
DNA Mediated by Triple Helix Formation.....	98
Chapter 4: Site Specific Cleavage of Human Chromosomal DNA	
Mediated by Triple Helix Formation.....	152
Huntington Disease: Review of the Mapping Literature	152
Part I: Identification of Triple Helix Target Sites Within	
Unsequenced Genetic Markers	164
Part II: Site Specific Cleavage of Human Chromosome 4	197
Chapter 5: Proposal: Identification of the Huntington Disease	
Gene by Analysis of Cleavage Products	231

Appendix A: Oligonucleotide Sequences	244
Appendix B: A Chemical Approach to Recognition and Cleavage of a Yeast Chromosome at a Single Site in High Yield.....	251

List of Tables and Figures

Chapter 1.

Tables

Table 1.1.	Restriction enzymes commonly used for genomic mapping	4
Table 1.2.	Calculation of sequences recognized by 3'3' alternate strand triple helix formation.....	21
Table 1.3	Calculation of total sequence recognized by oligonucleotide directed triple helix formation.....	23

Figures

Fig. 1.1.	Isomorphous T•AT and C+GC base triplets	9
Fig. 1.2.	Oligonucleotide directed triple helix formation	10
Fig. 1.3.	DMT-protected T*-phosphoramidite-triethylester.....	13
Fig. 1.4.	G•TA base triplet within the pyrimidine motif.....	15
Fig. 1.5.	Structure of D3 and D8 novel heterocycles	16
Fig. 1.6.	G•GC, A•AT, and T•AT base triplets within the purine motif..	19

Chapter 2.

Figures

Fig. 2.1.	Cleavage of secondary sites in bacteriophage λ DNA	40
Fig. 2.2.	Sequences of primary and secondary cleavage sites.....	42
Fig. 2.3.	Cleavage of λ DNA with CT and MeCBrU oligonucleotides	44
Fig. 2.4.	Competition reaction with T* and non-T* oligonucleotides	48
Fig. 2.5.	Synthetic scheme for T*-CPG.....	50
Fig. 2.6.	Triple helix formation with 5' and 3' T* oligonucleotide.....	52
Fig. 2.7.	Histogram of high resolution cleavage with 5' and 3' T* oligonucleotides.....	53

Fig. 2.8.	Double strand cleavage of bacteriophage λ DNA with multiple T* oligonucleotides.....	55
Fig. 2.9.	Improved cleavage efficiency by pH cycling.....	61
Chapter 3.		
Figures		
Fig. 3.1.	Map of primary and secondary cleavage sites of CTT* on chromosome III.....	88
Fig. 3.2.	Cleavage of yeast chromosome III as a function of oligonucleotide composition and pH.....	90
Fig. 3.3.	Cleavage of yeast chromosome III with nonspecific oligonucleotides.....	93
Fig. 3.4.	Map of cleavage sites recognized by nonspecific oligonucleotides.....	95
Fig. 3.5.	Map of pUCLEU2A showing triple helix target site.....	101
Fig. 3.6.	Enzymatic cleavage of pUCLEU2A mediated by triple helix formation.....	103
Fig. 3.7.	Construction of pUCLEU2 and its derivatives.....	137
Fig. 3.8.	Sequence of pUCLEU2 from Xho-235 to pUC19 ligation site.....	144
Chapter 4.		
Figures		
Fig. 4.1.	Physical map of interval from D4S10 to 4pter.....	156
Fig. 4.2.	Recombination events predicting a distal location for the HD gene.....	157
Fig. 4.3.	Recombination events predicting an internal location for the HD gene.....	158

Fig. 4.4.	Possible locations for the HD gene based on recombination events.....	160
Fig. 4.5.	Strategies for detecting triple helix binding sites that overlap methylase restriction enzyme sequences.....	165
Fig. 4.6.	Sequences of oligonucleotide primers for target site search by PCR.....	170
Fig. 4.7.	Orientation of PCR primers at hybridization sites within the pUCLEU2 plasmid series.....	171
Fig. 4.8.	PCR of the pUCLEU2 plasmid series using a partially degenerate primer for amplification.....	173
Fig. 4.9.	Detection of target site within 8C10I5 by PCR amplification.....	179
Fig. 4.10.	General scheme for identification of target sites by triple helix mediated enzymatic cleavage.....	182
Fig. 4.11.	Domains and sequences of oligonucleotides used for target site search by enzymatic cleavage.....	183
Fig. 4.12.	Oligonucleotide concentration as a function of sequence and sequence class.....	185
Fig. 4.13.	Site specific cleavage of pUCLEU2A using a partially degenerate oligonucleotide specific for EcoRI methylase.....	187
Fig. 4.14.	Detection of a AluI/triplex target site in 8C10I5	190
Fig. 4.15.	Subcloning and sequence determination of the AluI/triplex target site in cosmid clone 8C10I5.....	193
Fig. 4.16.	Map and sequence of AluI/triplex target site on human chromosome 4.....	194
Fig. 4.17.	Cleavage of human AluI/triplex target site in pUCLEU2E	200
Fig. 4.18.	Cleavage of human AluI/triplex target site in yeast genomic DNA.....	203

Fig. 4.19.	PCR analysis of D4S10 (ter) locus H5.52	211
Fig. 4.20	Cleavage of D4S113 human chromosome 4 target site in pUCLEU2P	213
Fig. 4.21.	G-rich target site from D4S113	215

Chapter 5

Figures

Fig. 5.1.	Pedigree of family1 from which matched chromosomes were derived	235
Fig. 5.2.	Scheme for identification of the HD mutation.....	238

Chapter 1

Introduction

Direct physical isolation of specific segments of DNA from the human genome is a major goal in human genetics. A number of technologies have recently been developed to manipulate and characterize increasingly large DNA fragments with the intention of generating detailed genetic and physical maps of the human genome (reviewed in 1-3). These technologies have revolutionized the way DNA is separated, cloned, and mapped. To complement these efforts, rare cleaving strategies are being developed to infrequently cleave large chromosomal DNA at specific sites (3).

Genetic and Physical Mapping of Human Chromosomal DNA. Mapping of human genetic disorders was originally limited to sex linked diseases or those resulting from cytogenetically detectable chromosomal aberrations (1, 4). Techniques utilizing linkage between restriction fragment length polymorphisms (RFLP) have facilitated detailed genetic mapping of human chromosomal DNA (5, 6). Polymorphisms such as single base changes that disrupt restriction enzyme cleavage sites or variations in the number of tandem repeats (reviewed in 7) have served as physical signals to test for the rate of recombination between two linked loci (8). A rough estimate of the physical distance between two markers can be determined by analyzing a number of recombination events to access the frequency at which two RFLP are segregated. The distances are measured in percent recombination or centimorgans. Through multipoint linkage analysis the relative order of a number of genetic markers on a

chromosome can be assigned. Disease genes have been cloned solely by their map position by analyzing genetic markers for linkage with genetic diseases (1, 9-11).

The rate of recombination along the length of a chromosome is not constant, resulting in genetic maps that grossly under and over estimate the distance between genetic markers (12). Physical maps, however, measure the distance between genetic markers in base pairs, providing an accurate measure of the physical linkage between RFLPs and disease genes.

Physical maps are generated by either a top-down or a bottom-up approach (reviewed in 3). In the bottom-up approach, chromosomal DNA is cloned into cosmid or yeast artificial chromosome (13) vectors to create a library of small (less than 45 kb) to medium (400 kb average) sized DNA inserts. Maps are generated by ordering the library into a single contig for each chromosome, identifying the order of clones in the library by detecting overlap between each clone. In the top-down approach the chromosomes are broken into a few large pieces that are characterized and ordered. These pieces are then cut into progressively smaller pieces which are again characterized and ordered. Physical mapping by the top-down approach was originally limited by separation of the cleavage products and the availability of rare cutting agents capable of generating a small number of large cleavage products. Strategies to eliminate both these limitations have now been developed.

Physical mapping of human chromosomal DNA was facilitated by the development of pulsed-field gel electrophoresis for the separation of large DNA fragments (reviewed in 14, 15). Conventional electrophoresis in a constant electric field was limited to resolution of DNA smaller than 50 kb in size. By periodically reversing the polarity of the current, the

resolution power of electrophoresis has improved to the point that megabase DNA fragments can be resolved (16, 17). Separation is based upon the reorientation of DNA in the electric field as a function of fragment size, field strength, and pulse time. Larger molecules require a longer amount of time to reorient to the altered current vector, and therefore migrate more slowly through the matrix.

Physical Mapping by Rare Cutting Restriction Enzymes. The development of pulsed-field gel electrophoresis enabled 50 kb to 5 Mb DNA to be resolved, spurring efforts to create long range physical maps of chromosomal DNA to complement the genetic maps that were being developed. The quality of physical maps was limited by the paucity of rare cutting cleaving agents capable of generating large chromosomal fragments. Common endonucleases cut with sufficient frequency that maps could not be extended more than a few thousand base pairs. However, 6 bp endonucleases with CpG sequences were found to cut less frequently than statistically predicted. These enzymes (Table 1.1) combined with the discovery of a few CpG recognition enzymes with 7-8 bp specificities (18, 19) have enabled physical maps to be completed for *E. coli* (4.5 Mb) (20) and some regions of human chromosomal DNA (12).

While powerful and simple to use, rare cutting endonucleases suffer from a number of limitations. Because they predominantly cut within CpG rich sites, a rare dimeric sequence often found at the 5' end of genes (21, 22), all the endonucleases tend to cut multiple times within clustered "CpG islands" (23). This produces fragments of very large and very small sizes, making it difficult to detect overlap between different digests, and resulting in significant gaps in physical maps. An additional problem with rare cutting restriction enzymes is that CpG sequences are

Enzyme	Recognition Sequence	Average Product Size (kb)
NotI	<u>GCGGCCGC</u>	1000
MluI	<u>ACGCGT</u>	500
NruI	<u>TCGCGA</u>	500
SgrAI	<u>CRCCGGYG</u>	500
PvuI	<u>CGATCG</u>	300
RsrII	<u>CGG(A/T)CCG</u>	200
SfiI	GGCCNNNNNGGCC	200
AatII	<u>GACGTC</u>	100
AsuII	<u>TTCGAA</u>	100
BssHII	<u>GCGCGC</u>	100
ClaI	<u>ATCGAT</u>	100
EagI	<u>CGGCCG</u>	100
FspI	<u>TGCGCA</u>	100
SacII	<u>CCGCGG</u>	100
SalI	<u>GTCGAC</u>	100
SmaI	<u>CCCGGG</u>	100
SnaBI	<u>TACGTA</u>	100
XhoI	<u>CTCGAG</u>	100
PacI	TTAATTAA	20

Table 1.1. The names, recognition sequences and average fragment sizes of restriction enzymes commonly used for genomic mapping. Note that with few exceptions the enzymes contain a CpG sequence (underlined) in their recognition site.

frequently methylated at the 5-position of cytosine in mammalian DNA (21, 22), and many of the restriction enzymes are sensitive to the methylation state of the recognition sequence. This causes the observed cleavage pattern to vary with the cell passage and the tissue from which the DNA was isolated, significantly complicating data interpretation. Additionally, only a few rare cutting enzymes have been discovered, limiting the sequences that can be targeted.

Alternate Strategies for Rare Cutting Reagents. DpnI Cleavage of Methylated DNA. A number of strategies have been developed to extend the recognition site size, reduce the cleavage frequency, and improve the

generalizability of sequence recognition. Initial efforts for creation of rare cutting reagents used combinations of methylases with DpnI, a restriction enzyme that only cuts the sequence 5'-G^{Me}ATC-3' in which adenines on both strands are methylated. The 8 bp sequence 5'-TCG^{Me}ATCGA-3' could be uniquely targeted by modifying each 5'-TCGA-3' half site with TaqI methylase (24). This generated a central DpnI recognition sequence for cleavage (underlined). Using a similar approach with different methylases, larger specificities were achieved. Ten bp sequences 5'-ATCG^{Me}ATCGAT-3' and 5'-GAAG^{Me}ATCTTC-3' were uniquely cut by DpnI using ClaI methylase (5'-ATCG^{Me}AT-3') and MboI methylase (5'-GGAG^{Me}A-3'), respectively (24, 25). Simultaneous methylation with both ClaI and MboI, resulted in cleavage at the previous two sites and at 5'-ATCG^{Me}ATCTTC-3'. Finally, a 12 base pair target site (5'-TCTAG^{Me}ATCTAGA-3') was achieved with XbaI methylase (5'-TCT^{Me}AGA-3') and DpnI endonuclease (26). The 2.9 Mb circular genome of *Staphylococcus aureus* was cut into two fragments using ClaI methylase and DpnI (27). A number of additional specificities could also be developed through the use of other methylases (24). The technique is limited, however, by DpnI cleavage at hemimethylated sites, resulting in partial digestion and significantly smaller specificities than predicted.

Protein Mediated Achilles Heel Cleavage. Modification of restriction enzyme specificity through the use of methylases formed the basis of a more robust rare cutting strategy, protein mediated Achilles heel cleavage (28, 29). A number of DNA binding proteins have recognition sites greater than 10 bp in size, but do not cleave the DNA. Specific cleavage can be directed to these sequences by conferring the specificity of the DNA binding protein upon a restriction enzyme. If the recognition

sequence of the DNA binding protein contains a restriction enzyme site, that site can be uniquely cut by forming the protein/DNA complex and methylating to completion with the cognate methylase. The single site bound by the protein remains unmodified. Following protein removal the site is uniquely susceptible to digestion by the cognate restriction enzyme. The 20 bp site 5'-AATTGTGGAGCGCTCACAATT-3' was targeted for cleavage using *lac* repressor, HhaI methylase (5'-GCGC-3'), and HaeII endonuclease (5'-AGCGCT-3'). The 17 bp target site 5'-TACCACTGG-CGGTGATA-3' was cut using λ repressor and the HphI cognate enzymes (28). This approach to site specific cleavage was successfully employed on plasmid DNA, and the genomes of *E. coli* and *Saccharomyces cerevisiae* (29).

Protein mediated Achilles heel cleavage formed the basis for general rare cutting techniques (30-34), but was itself limited by a paucity of target sites and compatible proteins. The large recognition sites of the technique were not amenable to standard protocols for physical mapping because cleavage was too infrequent. Statistically, the 20 bp recognition site is predicted to occur only once every 10^{12} bp, or once in a genome 300 times larger than that of man. Therefore, to cut even once within total human genomic DNA the target site must be physically inserted. Although this is a fairly straightforward task in bacterial and yeast genomes, it is much more difficult in the human genome. Nevertheless, it produced quantitative single site cleavage of large genomic DNA at engineered target sites and could be useful for physical mapping techniques that utilize random transposon insertion of the target site.

Very Rare Cutting Restriction Enzymes. Other rare cutting reagents have also been developed. Boehringer Mannheim has recently

announced the discovery of a restriction enzyme with 18 bp specificity. Although this might suffer the same problems as the Achilles heel technique, i.e., it is too specific and not generalizable, it could also be useful for transposon based mapping strategies. A similar application could be found for the intron encoded maturases that have specificities exceeding 15 bp (35, 36).

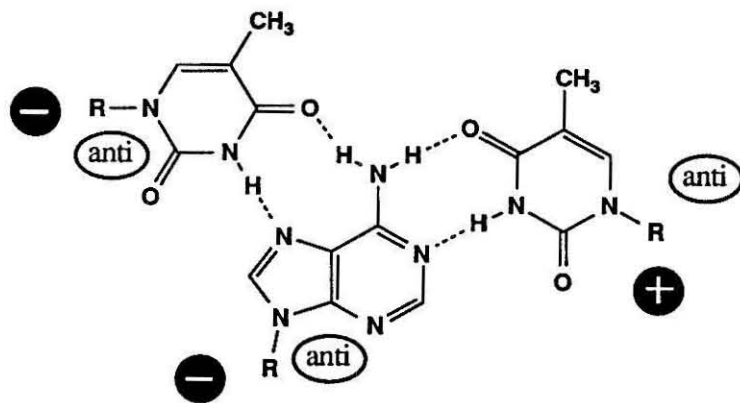
Radiation Hybrid Mapping. A technique based upon a significantly different cleaving strategy might also be considered a rare cutting reagent (37). Radiation hybrid mapping is a method used to randomly cut chromosomal DNA by high dose x-rays and recover broken chromosomal fragments in somatic cell hybrids. The frequency at which two markers are recovered together in the rodent-human hybrid clones is a direct measure of the distance between the markers. By estimating the frequency of breakage, a physical distance and order can be estimated. Although irradiation does not generate cleavage at a single region or target sequence, the recovery of single molecules in individual hybrid clones is equivalent to single site cleavage. The technique has been used to create a long range radiation map of human chromosome 21 spanning 20 Mb (37). The cell lines that result from the technique have proven useful for high frequency generation of region specific clones (38). Unlike the previous rare cutting strategies, radiation hybrid mapping does not rely on pulsed-field gel electrophoresis for resolution of the cleavage products, but it does facilitate the recovery and mapping of large regions of genomic DNA.

Microdissection. A crude technique to generate specific chromosomal fragments is chromosomal microdissection (39). Using either laser ablation or a high precision scalpel, physical regions of chromosomes marked by cytogenetic banding are cut out, pooled and

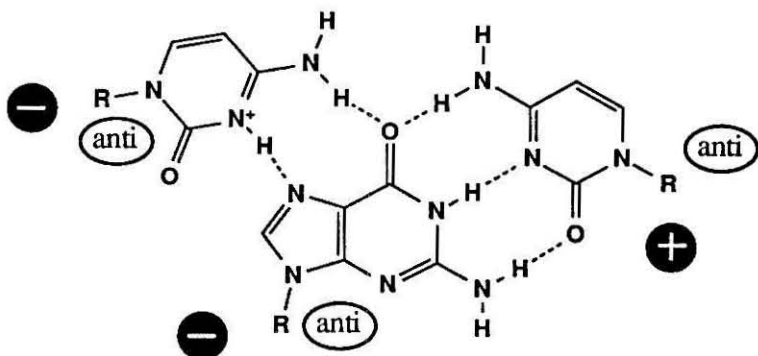
specific regions of the excised DNA amplified using the polymerase chain reaction. Although not satisfactory for many applications on genomic DNA, and not technically a mapping technique, it was able to generate flanking markers that lead to the successful isolation of the fragile site on the X-chromosome (40, 41).

Each of these techniques has been used for the mapping and isolation of genomic DNA. The effort would be significantly assisted, however, by the development of an efficient rare cutting strategy that could be specifically targeted to any of a large number of endogenous sequences throughout the human genome. Oligonucleotide directed triple helix formation is capable of unique molecular recognition of duplex DNA at binding sites greater than 12 base pairs in size, and is sufficiently general to recognize endogenous target sites at high frequency.

Oligonucleotide Directed Triple Helix Formation. Based upon work in polynucleotide mixing studies, Moser and Dervan demonstrated that pyrimidine oligodeoxyribonucleotides bind duplex DNA sequence specifically at homopurine sites to form a local triple helix structure (42). The pyrimidine oligomer is oriented in the major groove of DNA parallel to the Watson-Crick purine strand and binds by Hoogsteen hydrogen bonding. The "code" for triple helix specificity is derived from thymine (T) binding to adenine-thymine base pairs (T•AT base triplet) (43) and protonated cytosine (C+) binding to guanine-cytosine base pairs (C+GC base triplet) (Fig. 1.1) (44-47). By attaching the nonspecific affinity cleaving molecule EDTA•Fe(II) (48) to an oligonucleotide, Moser and Dervan demonstrated the binding of an oligonucleotide to a 15 bp purine target site. They found that the binding specificity was sufficient to define a single target site in plasmid DNA (4.0 kb) (Fig. 1.2) (42). As a function of



T•AT Triplet



C+GC Triplet

Fig. 1.1. Isomorphous T•AT and C+GC base triplets formed by Hoogsteen hydrogen bonding of a pyrimidine strand in the major groove of double helical DNA. The relative orientations of the three strands are indicated as + or -, and the bases are in the anti conformation.

the temperature of the binding reaction, the oligonucleotide could differentiate the target site from a similar site containing a single base mismatch by at least a factor of ten. They also demonstrated the importance of neutral to acidic pH and the requirement for a polycation such as spermine or cobalt-hexamine $\text{Co}(\text{NH}_3)_6^{3+}$ for efficient complex formation.

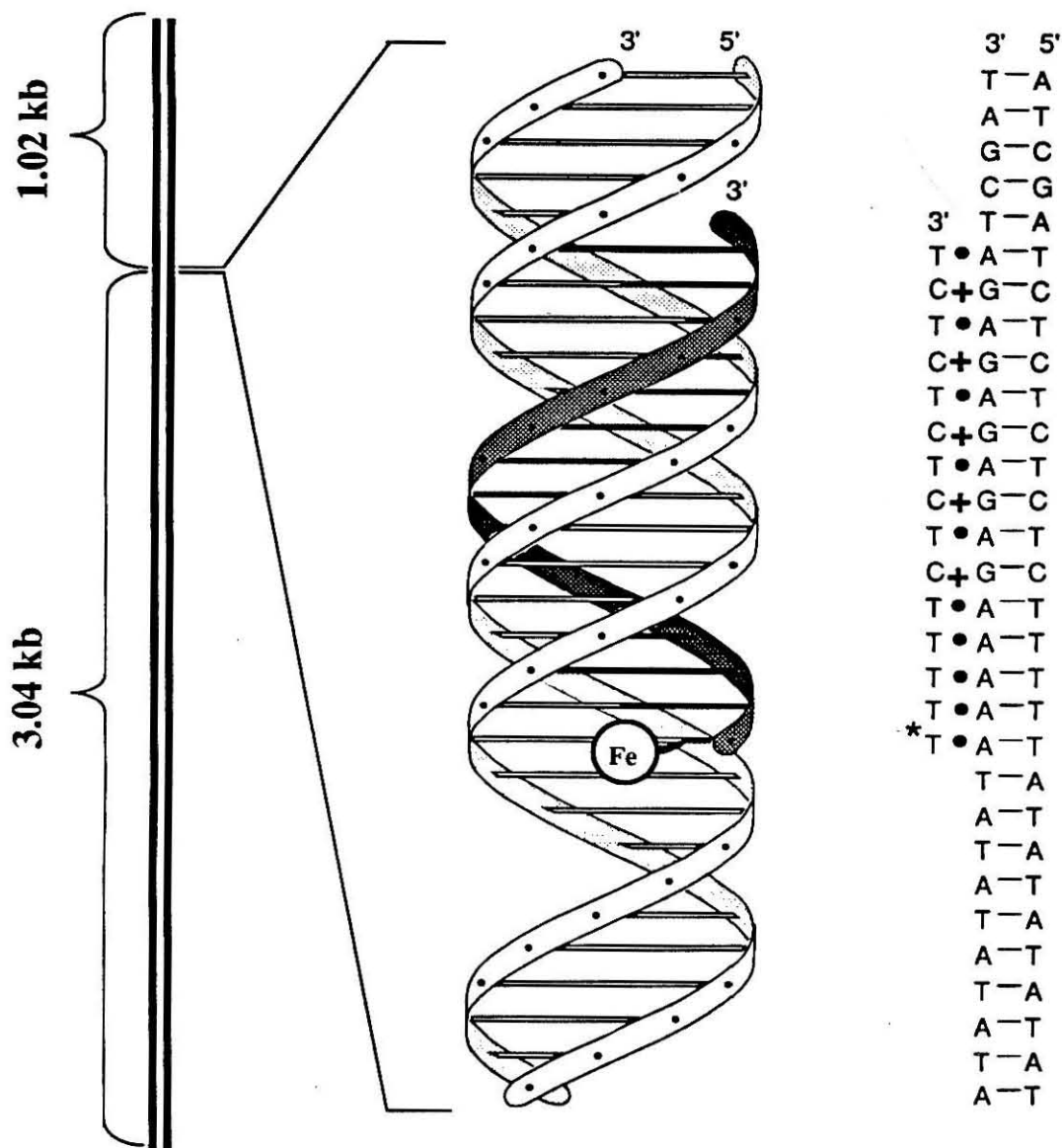


Fig. 1.2. Schematic of oligonucleotide directed triple helix formation showing the binding of a 15 base pyrimidine oligomer to a 15 bp homopurine target site. The third strand is bound in the major groove of the DNA duplex parallel to the purine strand. Moser and Dervan used this system to demonstrate site specific cleavage of plasmid DNA mediated by triple helix formation.

The structure of the triple helix was initially estimated by fiber diffraction of polynucleotide complexes (49, 50). It suggested that the triplex was in an A' conformation usually observed for double helical RNA under high salt conditions. While a high resolution image of the complex was not possible, the studies indicated some salient structural features including Hoogsteen hydrogen bonding for complexation. The bases of the helix were predominantly flat with a base tilt of $\approx 8.5^\circ$, and all three strands were in a C-3' endo sugar conformation with antiglycosidic bonds. The helix had a 30° rotation about the helical axis per residue, an axial rise per residue of 3.26 Å, and a pitch height of 39.1 Å.

Two dimensional ^1H NMR of an oligonucleotide triple helix complex provided additional insight into the structure of the triple helix (51-53). NMR directly observed the imino proton at the C3 position of cytosine in the third strand. The protonation of this position had been implicated by the marked pH sensitivity of the C+GC triplet. While both pyrimidine strands were found in the C-3' endo sugar conformation, the purine strand adopted a different conformation that appears to be sequence dependent (54-58).

Thermodynamics of the pyrimidine triple helix motif was studied using UV melting, NMR, CD, and differential scanning calorimetry (59, 60). Estimates for the enthalpic contributions (ΔH°) to triple helix formation at pH 6.5 and 200 mM NaCl indicated that the complex is stabilized by approximately 2.0 kcal/mol base triplet compared to 6.1 kcal/mol base pair for duplex DNA (59). Entropically the triplex is more favored than duplex formation, owing to the fewer degrees of freedom eliminated by the binding of a single stranded molecule to a stiff duplex.

The values are approximately 6.5 cal/mol·°K for the triplex transition compared to 18.5 cal/mol·°K for the duplex transition (59).

The ability of an oligonucleotide to sterically inhibit the activity of DNA binding proteins at a single target site was demonstrated by Maher *et al.* (30). These studies showed that a local triple helix complex can inhibit methylases, restriction endonucleases, and transcription factors when the triplex overlaps the protein binding site. Restriction enzyme inhibition facilitated a complete kinetic analysis of triple helix formation (61). Maher *et al.* demonstrated that the oligonucleotide binding reaction kinetics were pseudo-first order in oligonucleotide concentration with a comparatively slow on rate ($1.8 \times 10^3 \text{ L} \cdot (\text{mol of oligomer})^{-1} \cdot \text{s}^{-1}$) and an equilibrium dissociation constant of 10 nM. An increase in the pH of the binding reaction from pH 6.8 to 7.2 resulted in a small reduction in the association rate, but a significant increase in the dissociation rate (61). The specific inhibition of a transcription factor was subsequently studied *in vitro* for possible use of the triplex as an artificial gene-specific repressor (62).

Techniques for Oligonucleotide Mediated Cleavage of DNA.

Oligonucleotides were chemically modified to facilitate triple helix mediated cleavage of DNA. Moser and Dervan generated double strand cleavage at a single site in plasmid DNA (4.0 kb) by affinity cleaving using an EDTA•Fe(II) substitution at the C5 position of a single thymidine (T*) in the oligonucleotide (42, 48) (Fig. 1.3). The DNA backbone was cut by oxidation of the deoxyribose backbone by a short-lived nonspecific diffusible hydroxyl radical with double strand cleavage efficiencies of less than 10%. Perrouault *et al.* achieved photo-induced double strand

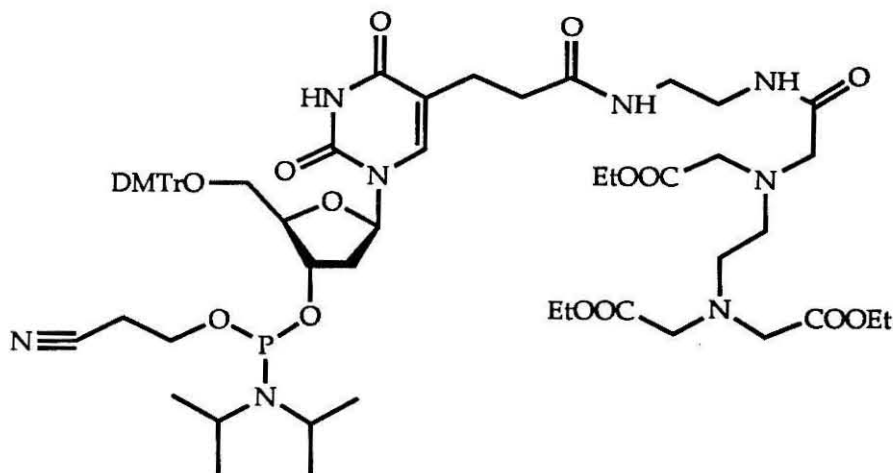


Fig. 1.3. DMT-protected T*-phosphoramidite-triethylester used for automated synthesis of affinity cleaving oligonucleotides containing EDTA.

cleavage of a 32-mer duplex by covalently attaching an ellipticine derivative to the 3' phosphate of the pyrimidine oligonucleotide and irradiating at wavelengths greater than 300 nm (63). Double strand cleavage efficiencies were not reported. Enzymatic cleaving functionalities were also used for triple helix mediated cleavage. Pei *et al.* synthesized a pyrimidine oligonucleotide conjugated at the 5' phosphate with staphylococcal nuclease and generated target specific double strand cleavage of plasmid DNA in greater than 75% yield (64). The cleavage occurred predominantly at A+T rich sites to the 5' side of the target sequence. The linker arm was sufficiently long, however, to accommodate cleavage at several base positions.

Cleavage by each of these oligonucleotide conjugates showed specificities mirroring that of the triple helix. Two additional approaches combined oligonucleotide binding specificity with a sequence specific cleaving function. Povsic and Dervan synthesized an oligonucleotide containing a bromo-acetyl moiety attached at the 5-position of the 5'

terminal thymidine (65). They observed sequence specific alkylation at the N7 position of G, two bases to the 5' side of the triple helix binding site (65). Presumably no alkylation would have occurred if a base other than G were present at this position. Work up with piperidine resulted in near quantitative cleavage of a single strand within the duplex (65). Double strand cleavage was achieved by positioning two single strand cleavage sites near each other, but on opposite strands (66). Alkylation of G was also achieved using a N4, N4-ethano-5-methyldeoxycytidine derivative, though its position within the binding site did not lead to additional specificity from the cleaving function (67).

Enzymatic cleavage of a target site could be achieved through the use of triple helix mediated Achilles heel cleavage. Maher *et al.* demonstrated that an oligonucleotide could prevent modification of an overlapping methylase recognition site (30). Following methylase inactivation and triplex disruption, the triple helix binding site was the only site susceptible to digestion by the cognate restriction enzyme. Cleavage specificity was the nonoverlapping sum of the oligonucleotide binding and the restriction enzyme specificities. Maher *et al.* used a 21 base oligonucleotide to mediate near quantitative double strand cleavage of a 23 bp target site using TaqI methylase and restriction enzyme (30). Using EcoRI methylase and endonuclease, Hanvey *et al.* cut a 15 bp target site with a 12 base oligonucleotide (31). Inhibition of methylase activity at a single site in plasmid DNA suggested that triple helix mediated Achilles heel cleavage could find application for the site specific cleavage of large genomic DNA. The combinatorial approach to target site recognition and cleavage in the later two strategies was critical to the achievement of single site cleavage in genomic DNA (32, 33, 66).

Recognition of Additional Sequences by Triple Helix Formation.

Oligonucleotides utilizing base triplets T•AT and C+GC can be designed to bind only a small fraction of all potential 16 bp target sites. Binding is limited by the lack of a recognition "code" for sequences containing pyrimidines within the purine strand, and the inability to recognize sequences containing contiguous G's. The number of sequences now targetable by pyrimidine oligonucleotide directed triple helix formation has been greatly extended by the discovery of additional triplet specificities, the synthesis of novel heterocycles, and synthetic modifications in the DNA backbone.

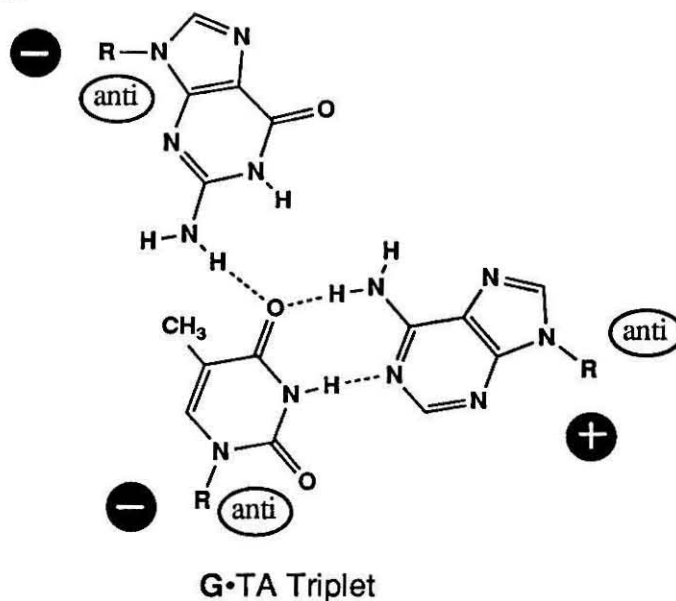
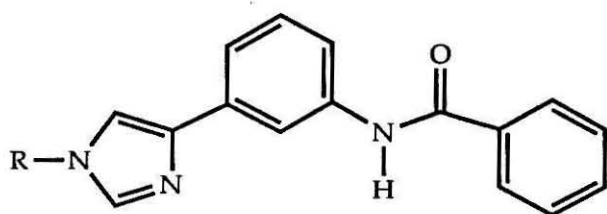


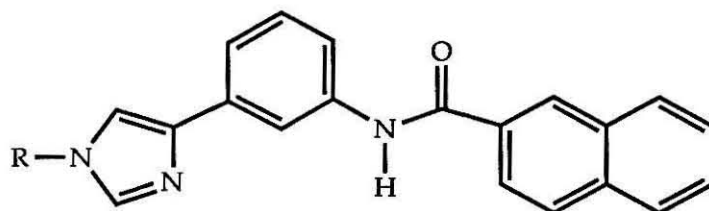
Fig. 1.4. Model proposed by Griffin and Dervan (68) for G recognition of a TA base pair within the pyrimidine motif.

The recognition of mixed purine/pyrimidine sequences was first achieved by the discovery of a specific interaction between guanine and a thymine-adenine base pair (G•TA base triplet) (68). Although the nature of the interaction awaits characterization by NMR, the simplest model for

the G•TA triplet places G in the anti conformation forming at least one hydrogen bond between G and T, likely N-2 of G with O-4 of T (68). Sites containing up to two TA bp were targetable by triplex formation using this novel triplet (68). Subsequent studies suggested that recognition might not be in the same plane as the rest of the base triplets resulting in a requirement for an AT base pair to the 5' side of the TA base pair (69). Although the affinity of G for the TA bp was low, it did show high specificity.



D₃: 3-benzamidophenylimidazole



D₈: 3-naphthamidophenylimidazole

Fig. 1.5. Chemical structure of D₃ and D₈ novel heterocycles for the recognition of TA and CG base pairs within the pyrimidine triple helix motif.

Novel heterocycles were also designed to recognize pyrimidine base pairs within purine rich sequences. The novel base 3-benzamidophenylimidazole (D₃) showed specificity for TA and CG base

pairs with a requirement for a AT base pair to the 3' side of the pyrimidine base (69, 70). Increasing the size of the phenyl ring to a naphthyl group (D_8) increased the binding affinity of the oligonucleotides high enough that mixed purine pyrimidine sites could be targeted at near neutral pH (71). Weak, but surprisingly specific binding, was achieved by incorporating an abasic sugar into the oligonucleotide (72). Although the binding affinity of the molecule was very low, it showed some specificity for pyrimidines at the base pair opposite the abasic position (72).

The recognition of G rich sequences in the pyrimidine motif was achieved by two novel heterocycles. Cytosine protonation destabilizes the C+GC triplet at slightly acidic or neutral pH (42). 2'-O-methyl-pseudoisocytidine and 3-methyl-5-amino-1H-pyrazolo-[4,3-d]pyrimidine-7-one (P_1) were each designed to include a proton donor for the N7 position of G (73, 74). Oligonucleotides containing either of these bases facilitated the efficient targeting of sequences containing a high G to A ratio at neutral or slightly basic pH.

Although the canonical base triplets only provided recognition of half the base pairs present in duplex DNA, it included a recognition "code" for the purine half of all base pair combinations. If the oligonucleotide could bind the purine half of each base pair by alternately binding to either face of the major groove, a generalizable solution to duplex recognition could be achieved. The parallel orientation of the third strand is such that binding to alternating faces of the groove requires the synthesis of oligonucleotides containing 3'3' and 5'5' linkages of the DNA backbone. Horne and Dervan described alternate strand triple helix formation using a 3'3' linkage that could target sequences of the class $(\text{purine})_m\text{N}_2(\text{pyrimidine})_n$ (75, 76). Ono *et al.* synthesized a 5'5' linkage

capable binding weakly to sequences of type (pyrimidine)_n(purine)_m (77). They also designed an oligonucleotide containing both a 5'5' and a 3'3' linkage capable of binding to alternate faces of the major groove by making two crossovers (77). Extension of this strategy could facilitate binding to a wide variety of target sites.

A Second Triple Helix Motif using Purine Oligonucleotides. In addition to a pyrimidine third strand, polynucleotide studies detected a three stranded complex containing a ratio of two purine strands per one pyrimidine strand. Hogan *et al.* reported transcription inhibition by a G rich oligonucleotide targeted to a promoter region of the *c-myc* gene parallel to the purine strand using putative G•GC, A•AT, T•TA, and C•CG triplets (78). Affinity cleaving experiments demonstrated that the actual orientation of the oligonucleotide was antiparallel to the purine strand and involved the formation of G•GC, A•AT, and T•AT base triplets (79) (Fig. 1.5). The complex was found to be pH insensitive because none of the triplets required protonation, but there was a requirement of polycation (79, 80). Beal and Dervan postulated that the most reasonable structure of the complex placed the phosphate backbone near the center of the major groove and the bases in the anti conformation (79). Although this structure has not been extensively characterized, thermodynamic parameters on a single sequence suggest this structure is significantly more stable than the pyrimidine triple helix, and has thermal stability and free energy of formation comparable to duplex formation (81).

Beal and Dervan have demonstrated that linear combinations of the purine and pyrimidine motifs are capable of binding to sequences of the type targeted by alternate strand triple helix formation (82). Because the binding orientation of the two triplex motif is reversed, there is no

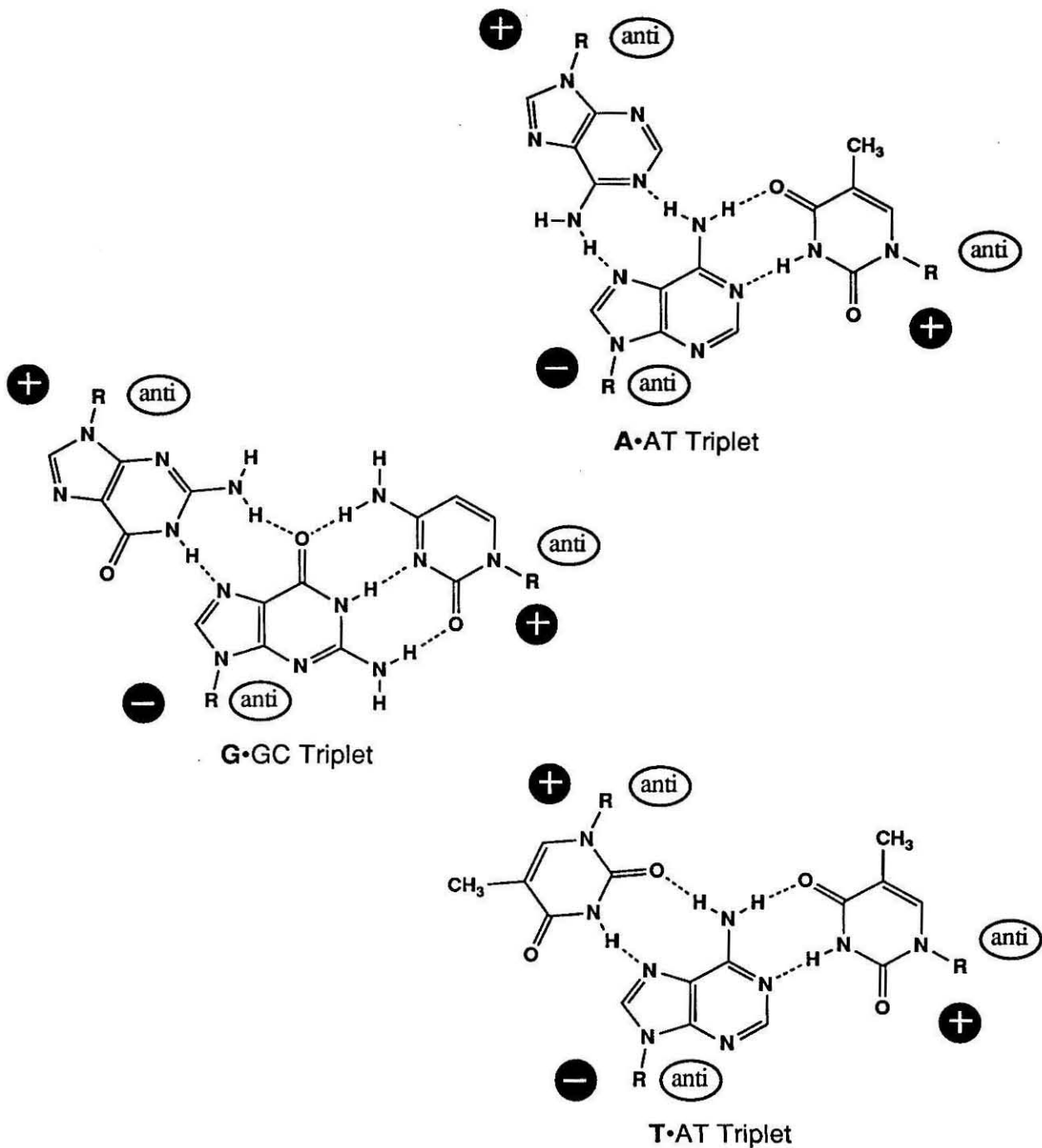


Fig. 1.6. Models proposed by Beal and Dervan (79) for G•GC, A•AT, and T•AT base triplets with a purine•purine-pyrimidine triple helix motif where the third strand is antiparallel to the purine Watson-Crick strand and the bases are in the anti conformation.

requirement for novel linkages at the crossover junctions.

Oligonucleotide Directed Triple Helix Formation Target Site Frequency. The binding site size of triple helix formation is sufficiently large to suggest that it might find application in the single site cleavage of genomic DNA, but sufficiently generalizable to suggest that it could be targeted to endogenous sequences within most genetic markers. The frequency of sequences targetable by triple helix formation is an important consideration in the possible utility of this technique.

A 16 bp recognition site ($n = 16$) defines 2,147,516,416 unique sequences (S) as determined from equations 3.1 and 3.2.

$$S = \frac{4^n}{2} \text{ (n is odd)} \quad (\text{Eq. 1.1}) \quad S = \frac{4^n}{2} + \frac{4^{n/2}}{2} \text{ (n is even)} \quad (\text{Eq. 1.2})$$

This number is equivalent to the size of the human genome ($\approx 3 \times 10^9$ bp) and indicates that a 16 bp recognition sequence is necessary to define a single site in human genomic DNA. Pyrimidine oligonucleotide directed triple helix formation utilizing T•AT and C+GC base triplets could recognize 65,536 (2^{16}) homopurine sequences (42). A second triplex motif utilizing G•GC, A•AT, and T•AT triplets statistically recognize the same subset of 65,536 sequences, but G-rich sequences would be recognized more efficiently with this motif (79, 80). The ability to recognize one TA base pair (sequence class $R_{15}T_1$) by the use of a single G•TA base triplet within the pyrimidine motif (68) increases the number of 16-bp sequences recognized by 524,288 ($16 \cdot 2^{15}$). Recognition of one CG base pair within the pyrimidine motif (sequence class $R_{15}C_1$) by the novel heterocycles D_3 or D_8

Sequence Class	N _A	N _B	Included Within Earlier Sequence Class:	# of New Sequences
R ₁ N _A N _B Y ₁₃	R	R	-	65,536
	R	Y	-	65,536
	Y	R	-	65,536
	Y	Y	-	65,536
R ₂ N _A N _B Y ₁₂	R	R	-	65,536
	R	Y	R ₁ N _A N _B Y ₁₃	0
	Y	R	-	65,536
R ₃ N _A N _B Y ₁₁	Y	Y	R ₁ N _A N _B Y ₁₃	0
	R	R	-	65,536
	R	Y	R ₂ N _A N _B Y ₁₂	0
R ₄ N _A N _B Y ₁₀	Y	R	-	65,536
	Y	Y	R ₁ N _A N _B Y ₁₃ & R ₂ N _A N _B Y ₁₂	0
	R	R	-	65,536
	R	Y	R ₃ N _A N _B Y ₁₁	0
R ₅ N _A N _B Y ₉	Y	R	-	65,536
	Y	Y	R ₂ N _A N _B Y ₁₂ & R ₃ N _A N _B Y ₁₁	0
	R	R	-	65,536
R ₆ N _A N _B Y ₈	R	Y	R ₄ N _A N _B Y ₁₀	0
	Y	R	-	65,536
	Y	Y	R ₃ N _A N _B Y ₁₁ & R ₄ N _A N _B Y ₁₀	0
	R	R	-	65,536
R ₇ N _A N _B Y ₇	R	Y	R ₅ N _A N _B Y ₉	0
	Y	R	-	65,536
	Y	Y	R ₄ N _A N _B Y ₁₀ & R ₅ N _A N _B Y ₉	0
	R	R	R ₅ N _A N _B Y ₉ & R ₆ N _A N _B Y ₈	0
	R	Y	R ₆ N _A N _B Y ₈	0
	Y	R	-	65,536
	Y	Y	R ₅ N _A N _B Y ₉ & R ₆ N _A N _B Y ₈	0
Total Sequences:				983,040

Table 1.2. Calculation of the number of sequences potentially recognized by 3'-3' alternate strand triple helix formation. The two base degenerate region in the center of the binding site, represented by N_A and N_B, complicates the calculation by causing significant sequence duplication when using simpler algorithm.

(70, 71) increases the number by an additional 524,288 ($16 \cdot 2^{15}$) sequences. These sites could also be recognized by a T•CG triplet within the purine motif (82). Binding to sites containing CG would not be unique. D₃ and D₈ also have affinity for TA base pairs in the pyrimidine motif (70, 71), and T has affinity for AT base pairs within the purine motif (82).

Alternate strand triple helix formation, in which junctions are incorporated into the oligonucleotide to facilitate binding to homopurine sequences that are alternatively on the Watson and the Crick strands (75-77), significantly increases the number of sites targetable by triplex formation. Oligonucleotides containing a 3'-3' linkage bind sequences of the type (purine)_nNN(pyrimidine)_{14-n} ($1 \leq n < 14$) affording 983,040 target sites (Table 1.2) (75, 76). Binding to these sites is not single sequence specific as the base pairs at the binding junction are not defined, i.e., a single 14 bp oligonucleotide with a 3'-3' junction can bind 16 different 16 bp sequences. Oligonucleotides with a 5'-5' linkage bind sequences of class (pyrimidine)_n(purine)_{16-n} ($1 \leq n < 16$) affording 524,288 single sequence specific target sites ($8 \cdot 2^{16}$) (77). These sequences can also be targeted by linear combinations of the purine and the pyrimidine motifs (82). While statistically this does not expand the number of recognizable sequences, it provides a second approach to binding the sites listed in this calculation.

The total number of sites recognized by triple helix formation is the sum of each sequence class less the sequence duplications. The calculation yields a total of 2,490,368 sites (Table 1.3). This represents 0.12% of all possible 16 bp sequences and suggests that a site targetable by triple helix formation will occur with an average frequency of once every 860 bp. The number of sequences recognized would increase exponentially if the motifs were combined, i.e., a double crossover, a crossover with a novel

Sequence Class	# of Sequences Recognized	# of Duplicated Sequences	# of Unique Sequences
R ₁₆	65,536	0	65,536
R ₁₅ T ₁	524,288	0	524,288
R ₁₅ C ₁	524,288	0	524,288
R _n (NN)Y _{14-n}	983,040	-65,536	917,504
Y _n R _{16-n}	524,288	-65,536	458,752
Total Sequences:			2,490,368

Table 1.3. List of five major sequence classes recognized by oligonucleotide directed triple helix formation. R and Y represent purine and pyrimidine, respectively (Column 1). 3'-3' and 5'-5' alternate strand triplex motifs duplicated a portion of the sequences recognized by novel base triplets requiring an adjustment in the totals (Column 3).

base triplet, two novel base triplets, etc. While sequence composition effects will significantly affect the validity of these calculations, the value is possibly an underestimate of the total because it does not include combinations of the sequence classes and homopurine sequences are observed to occur at greater than statistical frequency (83, 84). This calculation demonstrates that triple helix formation has become sufficiently generalized that targetable sites will occur in high frequency within human chromosomal DNA.

Description of Work. This thesis describes the use of oligonucleotide directed triple helix formation for the site specific cleavage of large genomic DNA. The work predominantly uses the pyrimidine triplex for target site recognition. Cleavage is achieved by affinity cleaving with oligonucleotides containing EDTA•Fe(II), double strand alkylation with oligonucleotides containing a bromo-acetyl electrophile, and Achilles heel cleavage with cognate methylases and restriction enzymes. The chapters are grouped according to the substrate of the cleavage reaction.

Chapter II describes the site specific cleavage of bacteriophage λ DNA by affinity cleaving. Strategies for optimization of cleavage yields are explored, and the specificity of triplex formation on a large, fully sequenced substrate is described as a function of oligonucleotide composition, pH and temperature. Cooperative binding of oligonucleotides to abutting target sites on bacteriophage λ DNA is demonstrated and the possible source of the cooperative interaction postulated.

Chapter III describes the site specific cleavage of an engineered target site on chromosome III (340 kb) of yeast genomic DNA (14 Mb) by triple helix mediated affinity cleaving and Achilles heel cleavage techniques. Secondary cleavage sites are mapped on the chromosomal substrate, nonspecific oligonucleotides are used to target endogenous sequences, and the specificity of oligonucleotides containing base substitutions is demonstrated. A complete protocol is included for insertion of triplex target sites into yeast chromosome III.

Chapter IV demonstrates triple helix mediated enzymatic cleavage of an endogenous target site near the tip of human chromosome IV. The 3.6 Mb fragment liberated by cleavage contains the entire candidate region for the Huntington disease gene. Cleavage of the target site is achieved in 80-90% yield. The chapter includes a review of the literature describing efforts to identify the HD mutation, and discusses the conflicting recombination data that places the mutation into either of two nonoverlapping regions. It also describes a number of protocols used to identify endogenous triple helix target sites that overlap methylation recognition sequences.

Chapter V suggests potential strategies using triple helix mediated cleavage that could assist in the search for the Huntington's disease mutation. Some preliminary experiments toward these goals are described.

Appendix A contains a complete listing of the oligonucleotides used for experiments in this thesis, and a brief description of their designed application.

Appendix B contains a manuscript prepared in collaboration with Tom Povsic, describing the quantitative site specific double strand cleavage of a yeast chromosome by alkylation. The paper analyzes the cleavage products from single base resolution on restriction fragments to the site specific cleavage of chromosomal DNA in yeast. The products are compatible with ligation into a restriction enzyme site with complementary ends. The cleavage specificity and yields of the chemical cleavage mechanism are comparable to that achieved by triple helix mediated enzymatic cleavage. The manuscript was submitted to the *Journal of the American Chemical Society*.

References

1. V. A. McKusick, Current Trends in Mapping Human Genes *FASEB J.* **5**, 12-20 (1991).
2. B. J. F. Rossiter, C. T. Caskey, Molecular Studies of Human Genetic Disease *FASEB J.* **5**, 21-27 (1991).
3. P. R. Billings, C. L. Smith, C. R. Cantor, New Techniques for Physical Mapping of the Human Genome *FASEB J.* **5**, 28-34 (1991).
4. R. P. Donahue, W. B. Bias, J. H. Renwick, V. A. McKusick, Probable Assignment of the Duffy Blood Group Locus to Chromosome 1 in Man *Proc. Natl. Acad. Sci. U.S.A.* **61**, 949-955 (1968).
5. D. Botstein, R. L. White, M. Skolnick, R. W. Davis, Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms *Am. J. Hum. Genet.* **32**, 314-331 (1980).
6. R. White, J. M. Lalouel, Chromosome Mapping with DNA Markers *Scient. Amer.* **258**, 40-48 (1988).
7. U. K. Landegren, R. Kaiser, C. T. Caskey, L. Hood, DNA Diagnostics-Molecular Techniques and Automation *Science* **242**, 229-237 (1988).
8. R. White, M. Leppert, D. T. Bishop, D. Barker, J. Berkowitz, C. Brown, P. Callahan, T. Hom, L. Jerominski, Construction of Linkage Maps with DNA Markers for Human Chromosomes *Nature* **313**, 101-105 (1985).
9. J. M. Rommens, M. C. Iannuzzi, B. S. Kerem, M. L. Drumm, G. Melmer, M. Dean, R. Rozmahel, *et al.*, Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping *Science* **245**, 1059-1065 (1989).
10. J. R. Riordan, J. M. Rommens, B. S. Kerem, R. Rozmahel, Z. Grzaelczak, J. Zielenski, *et al.*, Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA *Science* **245**, 1066-1073 (1989).
11. B. S. Kerem, J. M. Rommens, J. A. Buchanan, D. Markiewicz, T. K. Cox, A. Chakravarti, M. Buchwald, *et al.*, Identification of the Cystic Fibrosis Gene: Genetic Analysis *Science* **245**, 1073-1080 (1989).

12. J. C. Stephens, M. L. Cavanaugh, M. I. Gradie, M. L. Mador, K. K. Kidd, Mapping the Human Genome: Current Status *Science* **250**, 237-244 (1990).
13. D. T. Burke, G. F. Carle, M. V. Olson, Cloning of Large Segments of Exogenous DNA into Yeast by Means of Artificial Chromosome Vectors *Science* **236**, 806-812 (1987).
14. G. Chu, Pulsed-Field Gel Electrophoresis: Theory and Practice *Methods* **1**, 129-142 (1990).
15. E. Lai, B. W. Birren, S. M. Clark, M. I. Simon, L. Hood, Pulsed Field Gel Electrophoresis *BioTechniques* **7**, 34-42 (1989).
16. D. C. Schwartz, C. R. Cantor, Separation of Yeast Chromosome-Sized DNAs by Pulsed Field Gradient Gel Electrophoresis *Cell* **37**, 67-75 (1984).
17. S. M. Clark, E. Lai, B. W. Birren, L. Hood, A Novel Instrument for Separating Large DNA Molecules with Pulsed Homogeneous Electric Fields *Science* **241**, 1203-1205 (1988).
18. J. M. Nelson, S. M. Miceli, M. P. Lechevalier, R. J. Roberts, *FseI*, A New Type II Restriction Endonuclease that Recognizes the Octanucleotide Sequence 5'-GGCCGGCC-3' *Nuc. Acids Res.* **18**, 2061-2064 (1990).
19. N. Tautz, K. L. Kaluza, B. Frey, M. Jarsch, G. G. Schmitz, C. Kessler, *SgrAI*, A Novel Class-II Restriction Endonuclease from *Streptomyces griseus* Recognizing the Octanucleotide Sequence 5'-CR/CCGYA-3' *Nuc. Acids Res.* **18**, 3087 (1990).
20. Y. Kohara, K. Akiyama, K. Isono, The Physical Map of the Whole *E. coli* chromosome: Application of a New Strategy for Rapid Analysis and Sorting of a Large Genomic Library *Cell* **495-508**, (1987).
21. A. P. Bird, CpG-Rich Islands and the Function of DNA Methylation *Nature* **321**, 209-213 (1986).
22. A. P. Bird, CpG Islands as Gene Markers in the Vertebrate Nucleus *Trends Genet.* **3**, 342-347 (1987).
23. D. I. Smith, W. Golembieski, J. D. Gilbert, L. Kizyma, O. J. Miller, Overabundance of Rare-Cutting Restriction Endonuclease Sites in the Human Genome *Nuc. Acids Res.* **15**, 1173-1184 (1987).

24. M. McClelland, L. G. Kessler, M. Bittner, Site-Specific Cleavage of DNA at 8- and 10-Base Pair Sequences *Proc. Natl. Acad. Sci. U.S.A.* **81**, 983-987 (1984).
25. M. McClelland, Site-Specific Cleavage of DNA at 8-, 9-, and 10-bp Sequences *Method. Enz.* **155**, 22-32 (1987).
26. Y. Patel, E. Van Cott, G. G. Wilson, M. McClelland, Cleavage at the Twelve-Base-Pair Sequence 5'-TCTAGATCTAGA-3' using M⁺XbaI (TCTAGA) Methylation and DpnI (GA/TC) Cleavage *Nuc. Acids Res.* **18**, 1603-1607 (1990).
27. M. D. Weil, M. McClelland, Enzymatic Cleavage of a Bacterial Genome at a 10-Base Pair Recognition Site *Proc. Natl. Acad. Sci. U.S.A.* **86**, 51-55 (1989).
28. M. Koob, E. Grimes, W. Szybalski, Conferring Operator Specificity on Restriction Endonucleases *Science* **241**, 1084-1086 (1988).
29. M. Koob, W. Szybalski, Cleaving Yeast and *Escherichia coli* Genomes at a Single Site *Science* **250**, 271-273 (1990).
30. L. J. Maher, B. Wold, P. B. Dervan, Inhibition of DNA Binding Proteins by Oligonucleotide-Directed Triple Helix Formation *Science* **245**, 725-730 (1989).
31. J. C. Hanvey, M. Shimizu, D. Wells, Site-Specific Inhibition of EcoRI Restriction/Modification Enzymes by a DNA Triple Helix *Nuc. Acids Res.* **18**, 157-161 (1990).
32. S. A. Strobel, P. B. Dervan, Single-site Enzymatic Cleavage of Yeast Genomic DNA Mediated by Triple Helix Formation *Nature* **350**, 172-174 (1991).
33. S. A. Strobel, L. A. Doucette-Stamm, L. Riba, D. E. Housman, P. B. Dervan, Site-Specific Cleavage of Human Chromosome 4 Mediated by Triple Helix Formation *Science* **254**, 1639-1642 (1991).
34. L. J. Ferrin, R. D. Camerini-Otero, Selective Cleavage of Human DNA: RecA-Assisted Restriction Endonuclease (RARE) Cleavage *Science* **254**, 1494-1497 (1991).
35. A. Delahodde, V. Goguel, A. M. Becam, F. Creusot, J. Perea, J. Banroques, C. Jacq, Site-Specific DNA Endonuclease and RNA Maturase Activities of Two Homologous Intron-Encoded Proteins from Yeast Mitochondria *Cell* **56**, 431-441 (1989).

36. J. M. Wenzlau, R. J. Saldanha, R. A. Butow, P. S. Perlman, A Latent Intron-Encoded Maturase is Also an Endonuclease Needed for Intron Mobility *Cell* **56**, 421-430 (1989).
37. D. R. Cox, M. Burmeister, E. R. Price, S. Kim, R. M. Myers, Radiation Hybrid Mapping: A Somatic Cell Genetic Method for Constructing High-Resolution Maps of Mammalian Chromosomes *Science* **250**, 245-250 (1990).
38. C. A. Pritchard, D. Casher, E. Uglum, D. R. Cox, R. M. Myers, Isolation and Field-Inversion Gel Electrophoresis Analysis of DNA Markers Located Close to the Huntington Disease Gene *Genomics* **4**, 408-418 (1989).
39. H. J. Ludecke, G. Senger, U. Claussen, B. Horsthemke, Cloning Defined Regions of the Human Genome by Microdissection of Banded Chromosomes and Enzymatic Amplification *Nature* **338**, 348-350 (1989).
40. M. Djabali, C. Nguyen, I. Biunno, B. A. Oostra, M. G. Mattei, J. E. Ikeda, B. R. Jordan, Laser Microdissection of the Fragile-X Region: Identification of Cosmid Clones and of Conserved Sequences in this Region *Genomics* **10**, 1053-1060 (1991).
41. R. N. MacKinnon, M. C. Hirst, M. V. Bell, J. E. V. Watson, U. Claussen, H. J. Ludecke, G. Senger, B. Horsthemke, K. E. Davies, Microdissection of the Fragile-X Region *Am. J. Hum. Genet.* **47**, 181-187 (1990).
42. H. E. Moser, P. B. Dervan, Sequence-Specific Cleavage of Double Helical DNA by Triple Helix Formation *Science* **238**, 645-650 (1987).
43. G. Felsenfeld, D. R. Davies, A. Rich, Formation of a Three Stranded Polynucleotide Molecule *J. Amer. Chem. Soc.* **79**, 2023-2024 (1957).
44. M. N. Lipsett, The Interactions of Poly C and Guanine Trinucleotide *Biochem. Biophys. Res. Commun.* **11**, 224-228 (1963).
45. M. N. Lipsett, The Interactions of Poly C and Guanine Trinucleotide *J. Biol. Chem.* **239**, 1256-1260 (1964).
46. M. N. Lipsett, Complex Formation between Polycytidylic Acid and Guanine Oligonucleotides *J. Biol. Chem.* **239**, 1256-1260 (1964).

47. F. B. Howard, J. Frazier, M. N. Lipsett, H. T. Miles, Infrared Demonstration of Two and Three Strand Helix Formation Between Poly C and Guanosine Mononucleotides and Oligonucleotides *Biochem. Biophys. Res. Commun.* **17**, 93-102 (1964).
48. J. Dreyer, P. B. Dervan, Sequence-specific Cleavage of Single-stranded DNA: Oligodeoxynucleotide-EDTA•Fe(II) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 968-972 (1985).
49. S. Arnott, P. J. Bond, Structures for Poly(U)•Poly(A)•Poly(U) Triple Stranded Polynucleotides *Nature* **244**, 99-101 (1973).
50. S. Arnott, E. Selsing, Structures for the Polynucleotide Complexes Poly(dA)•Poly(dT) and Poly(dT)•Poly(dA)•Poly(dT) *J. Mol. Biol.* **88**, 509-521 (1974).
51. P. Rajagopal, J. Feigon, Triple-Strand Formation in the Homopurine:Homopyrimidine DNA Oligonucleotides d(G-A)₄ and d(T-C)₄ *Nature* **339**, 637-640 (1989).
52. P. Rajagopal, J. Feigon, NMR Studies of Triple-Strand Formation from the Homopurine-Homopyrimidine Deoxyribonucleotides d(GA)₄ and d(TC)₄ *Biochemistry* **28**, 7859-7870 (1989).
53. C. de los Santos, M. Rosen, D. Patel, NMR Studies of DNA (R+)_n•(Y)_n•(Y+)_n Triple Helices in Solution: Imino and Amino Proton Markers of T•A•T and C•G•C⁺ Base-Triple Formation *Biochemistry* **28**, 7282-7289 (1989).
54. M. M. W. Mooren, Polypurine/Polypyrimidine Hairpins Form a Triple Helix Structure at Low pH *Nuc. Acids Res.* **18**, 6523-6529 (1990).
55. J. M. L. Pieters, R. M. W. Mans, H. van den Elst, G. A. van der Marel, J. H. van Boom, C. Altona, Conformational and Thermodynamic Consequences of the Introduction of a Nick in Duplexed DNA Fragments: an NMR Study Augmented by Biochemical Experiments *Nuc. Acids Res.* **17**, 4551-4565 (1989).
56. D. S. Pilch, C. Levensen, R. H. Shafer, Structural Analysis of the (dA)₁₀•3(dT)₁₀ Triple Helix *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1942-1946 (1990).

57. K. Umemoto, M. H. Sarma, G. Gupta, J. Luo, R. H. Sarma, Structure and Stability of a DNA Triple Helix in Solution: NMR Studies on $d(T)_6 \cdot d(A)_6 \cdot d(T)_6$ and Its Complex with a Minor Groove Binding Drug *J. Amer. Chem. Soc.* **112**, 4539-4545 (1990).
58. V. Sklenar, J. Feigon, Formation of a Stable Triplex from a Single DNA Strand *Nature* **345**, 836-838 (1990).
59. G. E. Plum, Y. W. Park, S. F. Singleton, P. B. Dervan, Thermodynamic Characterization of the Stability and the Melting Behavior of a DNA Triplex: A Spectroscopic and Calorimetric Study *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9436-9440 (1990).
60. D. S. Pilch, R. Brousseau, R. H. Shafer, Thermodynamics of Triple Helix Formation: Spectrophotometric Studies on the $d(A)_{10} \cdot 2d(T)_{10}$ and $d(C_+3T_4C_+3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ Triple Helices *Nuc. Acids Res.* **18**, 5743-5750 (1990).
61. L. J. Maher, P. B. Dervan, B. Wold, Kinetic Analysis of Oligodeoxyribonucleotide-Directed Triple-Helix Formation on DNA *Biochemistry* **29**, 8820-8826 (1990).
62. L. J. Maher, P. B. Dervan, B. Wold, Analysis of Promoter-Specific Repression by Triple-Helical DNA Complexes in a Eukaryotic Cell-Free Transcription System *Biochemistry* (In Press).
63. L. Perrouault, U. Asseline, C. Rivalle, N. T. Thuong, E. Bisagni, C. Giovannangeli, T. Le Doan, C. Helene, Sequence-Specific Artificial Photo-Induced Endonucleases Based on Triple Helix-Forming Oligonucleotides *Nature* **344**, 358-360 (1990).
64. D. Pei, D. R. Corey, P. G. Schultz, Site-Specific Cleavage of Duplex DNA by a Semisynthetic Nuclease via Triple-helix Formation *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9858-9862 (1990).
65. T. J. Povsic, P. B. Dervan, Sequence-Specific Alkylation of Double-Helical DNA by Oligonucleotide-Directed Triple-Helix Formation *J. Amer. Chem. Soc.* **112**, 9428-9430 (1991).
66. T. J. Povsic, S. A. Strobel, P. B. Dervan, A Chemical Approach to Recognition and Cleavage of a Yeast Chromosome at a Single Site in High Yield. *J. Amer. Chem. Soc.* (In Preparation).
67. J. P. Shaw, J. F. Milligan, S. H. Krawczyk, M. Matteucci, Specific, High-Efficiency, Triple-Helix-Mediated Cross-Linking to Duplex DNA *J. Amer. Chem. Soc.* **113**, 7765-7766 (1991).

68. L. C. Griffin, P. B. Dervan, Recognition of Thymine•Adenine Base Pairs by Guanine in a Pyrimidine Triple Helix Motif *Science* **245**, 967-971 (1989).
69. L. L. Kiessling, P. B. Dervan, Effects of Sequence on the Stabilities of Pyrimidine Motif Triple Helices as Determined by Affinity Cleaving *Biochemistry* (In Press).
70. L. C. Griffin, L. L. Kiessling, P. Gillespie, P. B. Dervan, Recognition of All Four Base Pairs of Double Helical DNA by Triple Helix Formation: Design of Nonnatural Deoxyribonucleosides for Pyrimidine•Purine Base Pair Binding *In Preparation*
71. L. L. Kiessling, P. Gillespie, P. B. Dervan, (In Preparation).
72. D. Horne, P. B. Dervan, Effects of an Abasic Site on Triple Helix Formation Characterized by Affinity Cleaving *Nuc. Acids Res.* (In Press).
73. A. Ono, P. O. P. Ts'o, L. Kan, Triplex Formation of Oligonucleotides Containing 2'-O-Methylpseudoisocytidine in Substitution for 2'-Deoxycytidine *J. Amer. Chem. Soc.* **113**, 4032-4033 (1991).
74. J. S. Koh, P. B. Dervan, Design of a Nonnatural Deoxyribonucleoside for Recognition of GC Base Pairs by Oligonucleotide Directed Triple Helix Formation *J. Amer. Chem. Soc.* (In Press).
75. D. A. Horne, P. B. Dervan, Recognition of Mixed-Sequence Duplex DNA by Alternate-Strand Triple-Helix Formation *J. Amer. Chem. Soc.* **112**, 2435-2437 (1990).
76. S. McCurdy, C. Moulds, B. Froehler, Deoxyoligonucleotides with Inverted Polarity: Synthesis and Use in Triple-Helix Formation *Nucleosides & Nucleotides* **10**, 287-290 (1991).
77. A. Ono, C. N. Chen, L. S. Kan, DNA Triplex Formation of Oligonucleotide Analogues Consisting of Linker Groups and Octamer Segments that have Opposite Sugar-Phosphate Backbone Polarities *Biochemistry* **30**, 9914-9921 (1991).
78. M. Cooney, G. Czernuszewicz, E. H. Postel, S. J. Flint, M. E. Hogan, Site-Specific Oligonucleotide Binding Represses Transcription of the Human *c-myc* Gene in Vitro *Science* **241**, 456-459 (1988).

79. P. A. Beal, P. B. Dervan, Second Structural Motif for Recognition of DNA by Oligonucleotide-Directed Triple Helix Formation *Science* **251**, 1360-1363 (1991).
80. R. H. Durland, D. J. Kessler, S. Gunnell, M. Duvic, B. M. Pettett, M. E. Hogan, Binding of Triple Helix Forming Oligonucleotides to Sites in Gene Promoters *Biochemistry* **30**, 9246-9255 (1991).
81. D. S. Pilch, C. Levenson, R. H. Shafer, Structure, Stability, and Thermodynamics of a Short Intermolecular Purine-Purine-Pyrimidine Triple Helix *Biochemistry* **30**, 6081-6087 (1991).
82. P. A. Beal, Ph.D. Thesis (California Institute of Technology, 1994).
83. M. J. Behe, The DNA-Sequence of the Human Beta-Globin Region is Strongly Biased in Favor of Long Strings of Contiguous Purine or Pyrimidine Residues *Biochemistry* **1987**, 7870-7875 (1987).
84. A. M. Beatty, M. J. Behe, An Oligonucleotide Sequence Bias Occurs in Eukaryotic Viruses *Nuc. Acids Res.* **16**, 1517-1528 (1988).

Chapter II

Part I

Site Specific Cleavage of Bacteriophage Lambda Genomic DNA

The binding site size of oligonucleotide directed triple helix formation is sufficiently large to statistically define a single site in the human genome (1, 2). To achieve the goal of site specific cleavage of human genomic DNA, we examined triple helix formation on incrementally larger DNA substrates (3, 4). A pyrimidine oligonucleotide equipped with the affinity cleaving agent EDTA•Fe(II) (5) was initially shown to cut at a single site within an oligonucleotide duplex (10^1 bp), a restriction fragment (10^2 bp), and a prokaryotic plasmid (10^3 - 10^4 bp) (3). Bacteriophage λ DNA (48.5 kb) was chosen as a substrate in the 10^4 - 10^5 bp size range because it was fully sequenced and there were a number of potential endogenous target sites for affinity cleaving (6). Several issues were investigated using bacteriophage λ as the DNA substrate. These included: (i) studying the specificity of triple helix mediated cleavage of small genomic DNA (4), (ii) determining the cleavage efficiency of DNA embedded in agarose (4), (iii) assaying the specificity of oligonucleotides with novel base modifications, (iv) studying techniques for improved cleavage efficiency with affinity cleaving, and (v) investigating cooperative binding of oligonucleotides in a triple helical complex (7).

The site targeted in these studies was an 18 bp homopurine sequence located within the bacteriophage λ origin of replication, 39,138 bp from the left terminus (6, 8, 9). The target sequence, 5'-A₄GA₆GA₄GA-3', is essential for

replication of the bacteriophage genome (8-11), and is the initial DNA segment to become single stranded preceding replication (12-15). The sequence is located within the λ O protein coding sequence, immediately adjacent to four 19-bp sequence repeats of hyphenated dyad symmetry that are bound by the λ O protein during the initial steps of replication (12, 13, 16, 17). A number of sequence and structural characteristics of this region are consistent with formation of a transient intramolecular triple helical structure (H-form DNA) (18-21): (i) the sequence contains a strong purine bias that is a short homopurine-homopyrimidine mirror repeat (22, 23), (ii) a structural transition is observed that is dependent upon negative supercoiling (10), and (iii) the structure is partially S1 nuclease sensitive (14). Although there is currently no direct data to demonstrate that the structure adopted is H-form DNA, it is noteworthy that the target site chosen for triple helix formation in bacteriophage λ DNA is a phenotypically important endogenous DNA sequence, potentially capable of forming an intramolecular three stranded structure *in vivo*.

Reprinted from the *Journal of the American Chemical Society*, 1988, 110, 7927.
 Copyright © 1988 by the American Chemical Society and reprinted by permission of the copyright owner.

Double-Strand Cleavage of Genomic DNA at a Single Site by Triple-Helix Formation

Scott A. Strobel,[†] Heinz E. Moser,[‡] and Peter B. Dervan*

*Arnold and Mabel Beckman Laboratories of Chemical Synthesis, California Institute of Technology
 Pasadena, California 91125*

Received July 22, 1988

The sequence-specific cleavage of double helical DNA by restriction endonucleases is essential for many techniques in molecular biology, including gene isolation, DNA sequencing, and recombinant DNA manipulations.^{1,2} With the advent of pulsed-field gel electrophoresis, the separation of large segments of DNA is now possible.^{3,4} However, the recognition sequences of naturally occurring restriction enzymes are in the range of 4–8 base pairs, and hence their sequence specificities may be inadequate for isolating genes from large chromosomes (10^8 base pairs in size) or mapping genomic DNA.⁵

Pyrimidine oligonucleotides bind duplex DNA sequence specifically at homopurine sites to form a triple helix structure.^{6,7}

[†]Howard Hughes Medical Institute Doctoral Fellow.

[‡]Swiss National Foundation Postdoctoral Fellow.

(1) Smith, H. O. *Science (Washington, D. C.)* **1979**, *205*, 455.

(2) Modrich, P. *Crit. Rev. Biochem.* **1982**, *13*, 287.

(3) Schwartz, D.; Cantor, C. R. *Cell* **1984**, *37*, 67.

(4) Carle, G. F.; Frank, M.; Olson, M. V. *Science (Washington, D. C.)* **1986**, *232*, 65.

(5) (a) Dervan, P. B. *Science (Washington, D. C.)* **1986**, *232*, 464. (b) Dervan, P. B. In *Nucleic Acids and Molecular Biology (Vol. 2)*; Eckstein, F., Lilley, D. M. J., Eds.; Springer-Verlag: Heidelberg, 1988; pp 49–64.

(6) Moser, H. E.; Dervan, P. B. *Science (Washington, D. C.)* **1987**, *238*, 645.

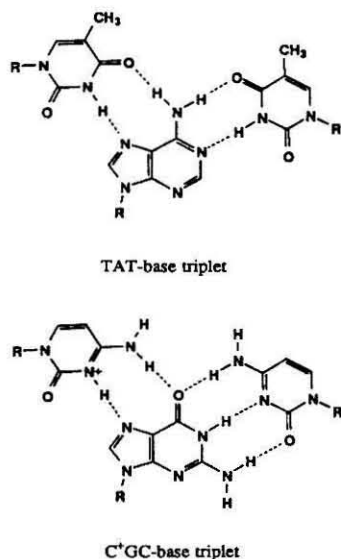


Figure 1. Isomorphous base triplets of TAT and C+GC. The pyrimidine oligonucleotide is bound by Hoogsteen hydrogen bonds in the major groove to the purine strand in the Watson-Crick duplex.

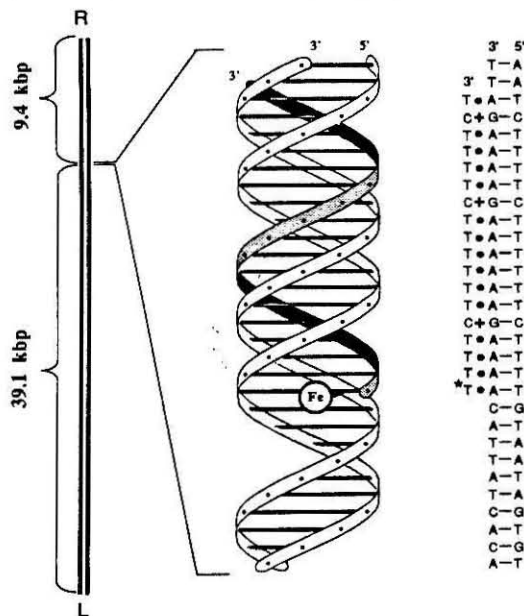


Figure 2. (Left) Double strand cleavage of λ DNA at the 18 base pair purine site affords two DNA fragments, 39.1 and 9.4 kbp in size. (Center) Simplified model of triple helix complex between the Watson-Crick homopurine-homopyrimidine site on λ DNA and the Hoogsteen bound oligonucleotide-EDTA-Fe.

The pyrimidine oligomer is oriented in the major groove of DNA parallel to the Watson-Crick purine strand by Hoogsteen base

(7) (a) Doan, T. L.; Ferrouault, L.; Praseuth, D.; Habhouh, N.; Decout, J.-L.; Thuong, N. T.; Lhomme, J.; Helene, C. *Nucl. Acids Res.* **1987**, *15*, 7749. (b) Praseuth, D.; Ferrouault, L.; Doan, T. L.; Chassignol, M.; Thuong, N.; Helene, C. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1349.

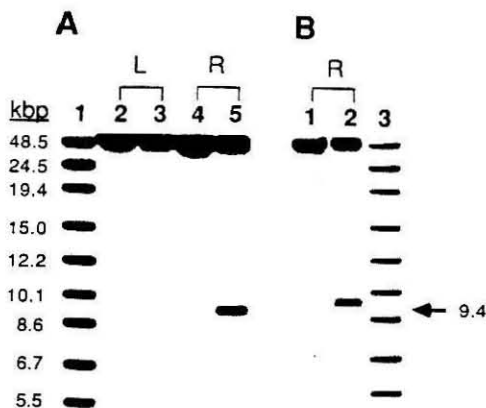


Figure 3. (A) Site-specific double strand cleavage of bacteriophage λ DNA. Autoradiogram of ^{32}P right (R) and left (L) end-labeled DNA, 48.5 kbp in length, on a 0.4% nondenaturing agarose gel. Lane 1, DNA size markers obtained by digestion of L end-labeled λ DNA with BamH I, Sma I, Apa I, SnaB I, and Xba I; digestion of right end-labeled DNA with BamH I, Sma I, and Xho I: 48.5 (undigested DNA); 24.5, 19.4, 15.0, 12.2, 10.1, 8.6, 6.7, and 5.5 kbp. Lane 2, L end-labeled intact λ DNA control. Lane 3, L end-labeled λ DNA with 0.8 μM oligonucleotide-EDTA-Fe. Lane 4, R end-labeled intact λ DNA control. Lane 5, R end-labeled λ DNA with 0.8 μM oligonucleotide-EDTA-Fe. (B) Site specific double strand cleavage of bacteriophage λ DNA in a 1% LMP agarose matrix. Autoradiogram of ^{32}P R end-labeled DNA on a 0.4% nondenaturing agarose gel. Lane 1, R end-labeled intact λ DNA control. Lane 2, R end-labeled λ DNA with 0.8 μM oligonucleotide-EDTA-Fe. Lane 3, DNA size markers as described above. Arrow on the right indicates 9.4 kbp cleavage fragment.

pairing (Figure 1).⁶⁻⁸ We have shown that oligonucleotides 15 bases in length, equipped with a cleaving function EDTA-Fe at one site in plasmid DNA which is 4.0 kilobase pairs (kbp) in size.⁶ Due to the length of the recognition site, in a formal sense, *this is 10⁶ times more sequence specific than restriction enzymes*.^{5,6} It is important to determine whether the full potential of this cleavage specificity can be realized in larger DNA. As an initial step beyond the cleavage of plasmid DNA, we report here the specificity and efficiency of double strand cleavage by an oligonucleotide-EDTA-Fe probe targeted to a naturally occurring 18 base pair sequence, 5'-A₄GA₄GA₄GA-3', that occurs once within the 48 502 base pairs (48.5 kbp) of the bacteriophage λ genome.^{9,10} The target sequence, located at base pair positions 39 138-56 from the left (L) end, is an essential purine cluster in λ 's origin of replication (ori) (Figure 2).^{11,12} Double strand cleavage at the 18 base pair target site would afford two DNA fragments, 39.1 and 9.4 kbp in size.

(8) For background on triple helical polynucleotides see: (a) Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1987**, *79*, 2023. (b) Michelson, A. M.; Massoulié, J.; Guschlbauer, W. *Prog. Nucleic Acids Res. Mol. Biol.* **1967**, *6*, 83. (c) Felsenfeld, G.; Miles, H. T. *Annu. Rev. Biochem.* **1967**, *36*, 407. (d) Lipsett, M. N. *Biochem. Biophys. Res. Commun.* **1963**, *11*, 224. (e) Lipsett, M. N. *J. Biol. Chem.* **1964**, *239*, 1256. (f) Howard, F. B.; Frazier, J.; Lipsett, M. N.; Miles, H. T. *Biochem. Biophys. Res. Commun.* **1964**, *17*, 93. (g) Miller, J. H.; Sobell, H. M. *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 1201. (h) Morgan, A. R.; Wells, R. D. *J. Mol. Biol.* **1968**, *37*, 63. (i) Lee, J. S.; Johnson, D. A.; Morgan, A. R. *Nucleic Acids Res.* **1979**, *6*, 3073. (j) Sanger, F.; Coulson, A. R.; Hong, G. F.; Hill, D. F.; Petersen, G. B. *J. Mol. Biol.* **1982**, *162*, 729-773.

(10) Strain c1857ind1Sam7 from New England Biolabs.

(11) Thompson, K. D.; Moore, D. D.; Kruger, K. E.; Furth, M. E.; Blattner, F. R. *Science (Washington, D. C.)* **1977**, *198*, 1051-1056.

(12) Furth, M. E.; Wickner, S. H. In *Lambda II*; Hendrix, R. W.; Roberts, J. W.; Stahl, F. W.; Weisberg, R. A., Eds.; Cold Spring Harbor, NY: 1983; pp 145-173.

Linearized bacteriophage λ DNA was labeled with ^{32}P individually at the left (L) and right (R) ends (3' fill-in reaction) with AMV reverse transcriptase. An 18 base homopyrimidine oligonucleotide, 5'-T* $\text{T}_3\text{CT}_6\text{CT}_4\text{CT}$ -3', with thymidine-EDTA (T*)¹³ at the 5' end, was synthesized by automated methods. The oligonucleotide-EDTA (0.8 μM) was mixed with $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ (1.0 μM) and spermine (1.0 mM). The oligonucleotide-EDTA-Fe(II)/spermine was added to a solution of L (or R) ^{32}P end-labeled λ DNA (approximately 4 μM in base pairs) in 100 mM NaCl and 25 mM tris-acetate at pH 7.0.¹⁴ After 0.5 h incubation at 24 °C, 4 mM dithiothreitol was added to initiate strand cleavage. The reaction was allowed to continue for 6 h (24 °C). The double strand cleavage products were separated on a 0.4% vertical agarose gel which resolves DNA segments up to 25 kbp in size (Figure 3). Therefore, DNA uniquely labeled at the R and L ends allows cleavage site analysis of the entire 48.5 kbp of the λ genome.

No cleavage was observed for the first 25 kbp of the L end-labeled DNA (Figure 3A, lane 3). However, a single major cleavage site 9.4 kbp from the right (R) end of the 48.5 kbp DNA was visualized on the autoradiogram (Figure 3A, lane 5).¹⁵ This was quantitated by scintillation counting of the individual bands. The oligonucleotide-EDTA-Fe probe afforded double strand cleavage of λ DNA at the target sequence, 5'-A₄G₆G₄G₄G₄-3', with an efficiency of 25% (Figure 3).¹⁶ Within the limits of our detection, overexposure of the autoradiogram revealed no secondary cleavage sites. We estimate that all secondary sites were cleaved at least 30-fold less efficiently than the primary sequence.¹⁷

In order to prevent random shearing of the DNA caused by pipetting or vortexing,^{3,4} large DNA is routinely manipulated by embedding it in low melting point (LMP) agarose and diffusing reagents into the matrix. To test whether triple helix mediated site-specific cleavage can occur with oligonucleotide-EDTA-Fe within an agarose matrix, end-labeled λ DNA was embedded in 1% LMP agarose gel, and the cleavage reactions were performed as described above (24 °C, pH 7.0, 6 h, initial incubation time was extended to 1 h to allow for complete diffusion into the matrix)

followed by electrophoresis. Primary site-specific double strand cleavage occurred in the agarose matrix with an efficiency of 25% (Figure 3B, lane 2).

In conclusion, this work has implications for human genetics. The way is now clear for the development of a triple helix strategy to isolate large segments of genomic DNA for mapping and sequencing.

Acknowledgment. This work was supported in part by the National Institutes of Health (GM 35724), the Caltech Consortium in Chemistry and Chemical Engineering (Founding members: E. I. duPont de Nemours and Company, Inc., Eastman Kodak Company, Minnesota Mining and Manufacturing Company, and Shell Oil Company Foundation), and a National Research Service Award (T32GM07616) from the NIH-GM :0 S.A.S.

(13) Dreyer, G. B.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 968.

(14) Concentrations given are final dilutions.

(15) The complementary fragment, 39.1 kbp from the L terminus, was not resolved from the intact λ DNA but can be visualized in underexposed autoradiograms as a shoulder with intensity equal to the 9.4 kbp fragment.

(16) 25% efficiency means that one-quarter of all the DNA underwent double strand cleavage at one site.

(17) Although the above conditions are optimal for single-site cleavage, reaction parameters can be modified for the purposes of searching large DNA for sequences of partial homology with the primary target site. For example, identical reaction conditions performed at the reduced temperature of 0 °C for 24 h resulted in three minor sites of cleavage (25-fold less efficient) which map to 22.4, 37.9, and 47.7 kbp from the L terminus. Sequences partially homologous with the primary site are found at each of these positions. One site, 5'-AACAA₄AAAAAAG-3', contains two mismatches in the first 12 nucleotides from the 5' end of the oligonucleotide-EDTA-Fe probe. Two other sites, 5'-AAAAGAAAAATGAA-3' and 5'-AAATGAAATAAAGAA-3', contain one and two mismatches in the first 14 base pairs, respectively. All other sequences in λ DNA with a similar degree of sequence homology (two mismatches in 12) were not cleaved within the limits of our detection and have in common the mismatch at or adjacent to the T* which contains the EDTA-Fe(II) cleaving moiety.

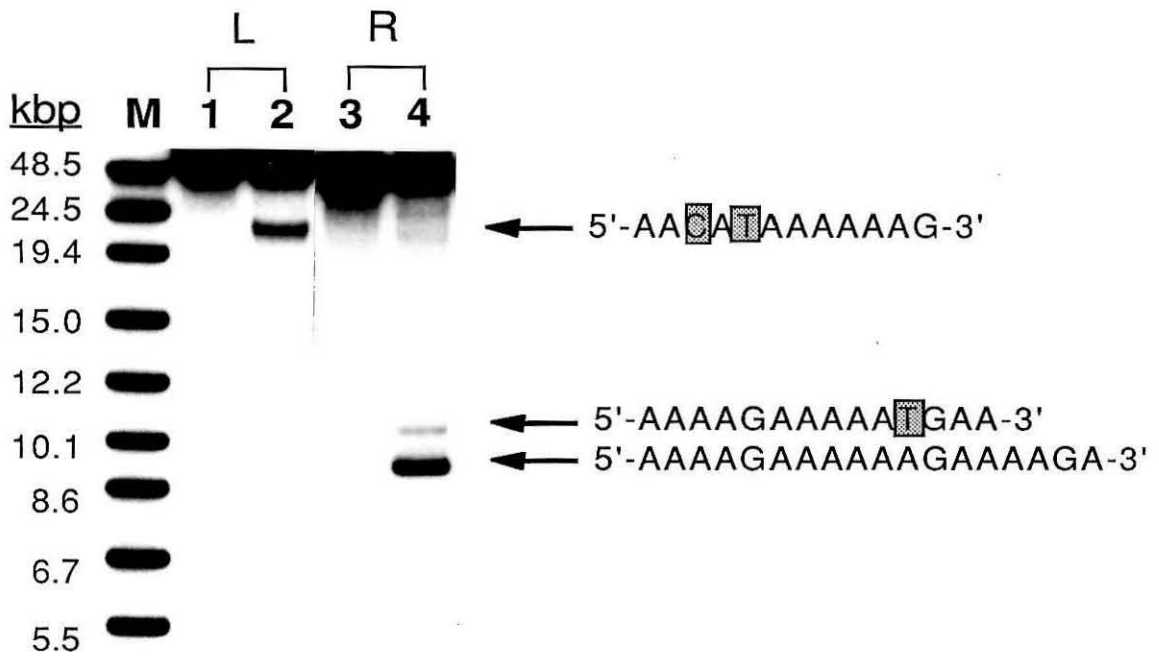
Specificity of Triple Helix Formation. Although the reaction conditions described were optimized for single site cleavage (4), the specificity of triple helix formation could be modified by adjusting the reaction conditions (3). By reducing either the temperature or the pH, sites of near, but imperfect sequence identity with the target site were cleaved. Reactions performed at 0° C pH 6.6 instead of 24° C pH 7.0 revealed three other cleavage sites that mapped 22.4, 37.9, and 47.7 kb from the left terminus of the bacteriophage λ genome (Fig. 2.1). Reactions performed at pH 7.0 and 0° C produced only one secondary cleavage site at 37.9 kb. This suggests that the sequence at 37.9 kb is more extensively homologous with the target site than the other two sites detected under less stringent conditions.

Sequences at the physical map positions of the observed cleavage products contained partial homology with the target site (6). A sequence containing only one mismatch within 14 contiguous base pairs (beginning with an A recognized by T* at the 5' end of the purine sequence) was identified at position 37.9 kb (Fig. 2.2). As predicted from the affinity cleaving data, no other sequences with 13 of 14 base pair homology were present in the bacteriophage λ genome. The sequences mapping to 22.4 and 47.7 kb were less homologous, containing two mismatches within 12 and 14 contiguous base pairs, respectively (Fig. 2.2). Unlike the 37.9 kb product, sequences with homology similar to the 22.4 and 47.7 sites were present, but none of these sites were cleaved within the limits of detection (>2% cleavage). Each of the partially homologous sequences not cut by the oligonucleotide contained a mismatch at or adjacent to the base pair bound by T* (Fig. 2.2). Presumably, the mismatch near T* generates a local structural distortion that effects the cleavage efficiency of the EDTA•Fe(II). A similar homology requirement was observed using oligonucleotides targeted to sites within adenovirus-2 DNA.

Fig. 2.1. Cleavage of secondary sites in bacteriophage λ DNA by oligonucleotide-EDTA•Fe(II). Autoradiogram of ^{32}P right (R) and left (L) end-labeled DNA, 48.5 kb in length, reacted with oligonucleotide-EDTA•Fe(II) at pH 6.6, 0° C. Reactions were performed on R or L end-labeled λ DNA (4 μM in base pairs), in 100 mM NaCl, 25 mM tris-HCl, 0.8 μM oligonucleotide-EDTA, 1.0 μM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$, 1.0 mM spermine•4HCl. Oligonucleotide, $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2\cdot 6\text{H}_2\text{O}$ and spermine were preequilibrated together for 5 minutes and then added to the buffered salt solution of DNA. Triple helix formation was allowed to proceed for 30 min. at 0° C after which DTT was added to a final concentration of 4 mM. The cleavage reaction was allowed to proceed for 24 hours. Cleavage products were resolved on a 0.4%, 1xTBE agarose gel. The gel was dried under vacuum and an autoradiogram obtained with an intensifying screen at -70° C.

Lane M, DNA markers of the size indicated. Lane 1, L end-labeled intact λ DNA control. Lane 2, L end-labeled λ DNA with oligonucleotide-EDTA•Fe(II). Lane 3, R end-labeled intact λ DNA control. Lane 4, R end-labeled λ DNA with oligonucleotide-EDTA•Fe(II). Cleavage products at R 9.4 kb (base position 39.1), R 10.8 kb (base position 39.7), R 26.1 kb and L 22.4 kb (base position 22.4) are indicated by arrows. The sequences in the region of cleavage most homologous with the target site are also indicated. Mismatches between the target site and the sequence identified at this locus are marked by a shaded box. The intensity of the 22.4 kb product is unusually high. The intensity is typically equivalent to the complementary 26.1 kb product seen in lane 4. The secondary target sites were cut 3 to 10 fold less efficiently than the primary target site.

Identical cleavage reactions also revealed a 0.8 kb product with R end-labeled DNA (base position 47.7) when separated on a 0.8% agarose gel for 100 V·hr (data not shown). The homologous sequence identified at this position was 5'-AAATGAA-TAAAGAA-3', where the mismatches are underlined. The cleavage efficiency was 3-fold less intense than the 9.4 kb major product. The 0.8 kb band had migrated off the bottom of the 0.4% gel shown here.



A	Base Position	5'- <u>T</u> [*] TTTTCTTTTTTTCTTTTCT-3'	Homology
	39,138	5'-T AC <u>AAAAGAAAAAAGAAAAGATTA</u> -3'	18/18
B	37,867	5'-AAAA <u>AAAGAAAAATGAACTTGGCT</u> -3'	13/14
	47,736	5'-AC <u>AAAATGAATAAAGAACAATCTG</u> -3'	12/14
	22,365	5'-AG <u>AAACAAAAAAAAGCCTGATGCA</u> -3'	10/12
C	36,829	5'-CAAC <u>CAAACAAAAAAGATGGGAATC</u> -3'	11/13
	43,336	5'-TCG <u>ACAACAAAAAAGACCTGCTTA</u> -3'	11/13
	36,946	5'-TGG <u>TAAAGAGAAAAGTTTTTCCAT</u> -3'	10/12
	30,664	5'-CGA <u>ATAATAAAAAAGGAGCCTGTA</u> -3'	10/12

Fig. 2.2. Sequences of primary and secondary cleavage sites within bacteriophage λ genomic DNA. **A.** Sequence of the oligonucleotide used for triple helix mediated affinity cleaving, and its primary target site. **B.** Secondary cleavage sites recognized and cut by oligonucleotide at low efficiency. **C.** Other potential secondary cleavage sites within bacteriophage λ DNA that have comparable homology to the target sequence as those listed in **B**, but were not cut by oligonucleotide-EDTA • Fe(II). These sequences have in common a mismatch at or adjacent to T*.

The base position in the bacteriophage λ genome is indicated to the left. The homologous region is underlined. The degree of homology with the target site is listed to the right. Mismatched bases within the underlined region are shown in outline. All sequences represent the maximum alignment of the oligonucleotide-EDTA • Fe(II) within the region where cleavage is observed.

Specificity of Oligonucleotides Containing MeC and BrU substitutions.

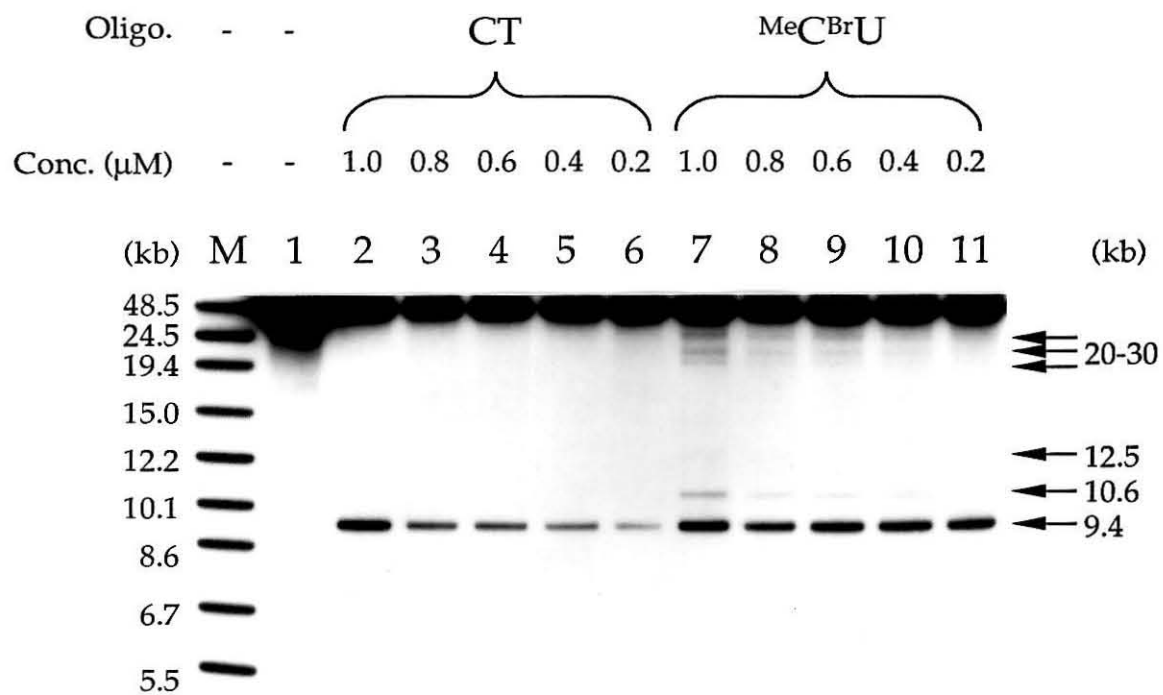
Povsic and Dervan reported that replacement of C with 5-methylcytosine (MeC) increased the affinity of the oligonucleotide for the target site and extended the pH range for binding (24). They also reported that substitution of T with 5-bromouracil (BrU) increased the binding affinity but did not change the pH profile of binding (24). Because the assay was performed on a plasmid with a limited number of partially homologous secondary binding sites, it was important to determine if these base substitutions affected the specificity of triple helix formation as assayed on larger DNA substrates.

An oligonucleotide specific for the bacteriophage λ target site was synthesized with T* at the 5' terminus and MeC and BrU substitutions for C and T, respectively. The specificity of this modified oligonucleotide was directly compared to the original CT oligomer (Fig. 2.3). At pH 7.0, 24° C several secondary cleavage sites were detected with the MeCBrU oligonucleotide that were not seen with the CT oligomer under similar cleavage conditions. These included sites at 37.9 and 22.4 kb (measured from the left terminus) that had previously been observed only at 0° C and pH 6.6. Additionally, there was a new product at map position 36 kb (\approx 12 kb fragment) and two other cleavage sites that mapped to positions between 20 and 30 kb from the left terminus. None of these products had previously been detected with the CT oligonucleotide. Cleavage at the secondary cleavage sites was not detected at lower oligonucleotide concentrations (<200 nM), although the MeCBrU oligonucleotide consistently cleaved more efficiently at lower concentration than the CT oligomer (Fig. 2.3).

Efforts to identify the new sequences recognized by the MeCBrU oligonucleotide were unsuccessful. No significant homology with the target site was detected at any of the new map positions, and no pattern could be

Fig. 2.3. Cleavage of bacteriophage λ DNA with CT and ^{32}P MeCBrU oligonucleotides as a function of oligonucleotide concentration.. Autoradiogram of ^{32}P right (R) end-labeled λ DNA (4 μM in base pairs) reacted in 100 mM NaCl, 25 mM tris-HCl, 1.0 mM spermine \cdot 4HCl, and 4 mM DTT at pH 7.0, 24 $^\circ$ C. CT and ^{32}P MeCBrU oligonucleotide-EDTA concentrations were varied between 200 nM and 1 μM using one molar equivalent of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. Oligonucleotide, $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2\cdot 6\text{H}_2\text{O}$ and spermine were preequilibrated together for 5 minutes and added to the buffered salt solution of DNA. The DNA was incubated with oligonucleotide-EDTA \cdot Fe(II) for 30 min. to facilitate triple helix formation after which DTT was added to a final concentration of 4 mM. The cleavage reaction was allowed to proceed for 24 hours. Cleavage products were resolved on a 0.4%, 1xTBE agarose gel. The gel was dried under vacuum and an autoradiogram obtained with an intensifying screen at -70 $^\circ$ C.

Lane M, DNA markers of the size indicated. Lane 1, R end-labeled intact λ DNA control. Lanes 2-6, reactions with CT oligonucleotide-EDTA \cdot Fe(II) at 1000, 800, 600, 400, and 200 nM, respectively. Lane 7-11, reactions with ^{32}P MeCBrU oligonucleotide-EDTA \cdot Fe(II) at 1000, 800, 600, 400, and 200 nM, respectively.



detected to explain why these sites were cleaved at the exclusion of other sites with similarly low homology. The data suggest that in addition to having higher affinity for the target and homologous secondary sites, the base substitutions produced altered specificity. Subsequent experiments on yeast genomic DNA demonstrated that the altered specificity was caused by BrU substitution and not from MeC (25). BrU is known to be mutagenic when incorporated into DNA, resulting in increased base misincorporation during replication and transcription (26-28). This could be caused by several mechanisms including BrU base ionization (29, 30) or tautomerization of the base to the enol form (31, 32). At neutral pH, greater than 40-fold more of the BrU residues are calculated to be ionized than T (7.4% compared to 0.16%) (30), and the enol tautomer of BrU is ten times more favored than that of T (32). The resulting alteration in hydrogen bonding properties could contribute to the broader sequence tolerance of BrU oligonucleotides, though the actual specificity of BrU within the triple helix has not been determined.

Improved Cleavage Efficiency. Although the target site in bacteriophage λ was cut more efficiently than any other sequence targeted by oligonucleotide directed affinity cleaving, the cleavage efficiency reached a maximum at 25% (4). Longer incubation times, higher oligonucleotide concentrations, additional iron and more reducing agent failed to improve the cleavage efficiency. In most cases, additional reagents and reaction time only increased the nonspecific nicking of the DNA substrate. To increase the efficiency of double strand cleavage, we employed two strategies: (i) cutting with oligonucleotides containing multiple T* moieties, and (ii) cutting the target site several times by pH cycling and fresh reagent addition.

Cleavage of Bacteriophage λ with Oligonucleotides Containing Multiple T Moieties.* A number of possible explanations could be advanced

to explain the cleavage maxima observed with affinity cleaving using oligonucleotide directed triple helix formation. These include: (i) incomplete occupation of the target site by the oligonucleotide, (ii) rapid probe exchange at the target site, (iii) low cleavage efficiency of EDTA•Fe(II) in the major groove, and/or (iv) autocleavage of the oligonucleotide.

Saturation and oligonucleotide exchange at the target site are not significantly limiting factors for cleavage. Competition experiments between probes with and without T* residues demonstrated that there is little probe exchange at the target site within the time course of the reaction, suggestive of a slow oligonucleotide off rate (Fig. 2.4). Footprinting and enzyme inhibition studies demonstrated that under the reported conditions, nearly all the target sites are occupied (33, 34). Subsequent kinetic analysis of triple helix formation demonstrated that the oligonucleotide off rate is several hours (35).

Quantitative cleavage of duplex DNA is observed with affinity cleaving agents that bind in the minor groove (1). Double strand oxidative cleavage is slowed, however, by low EDTA•Fe(II) reactivity in the major groove of DNA (36, 37). It is proposed that the oxidative species must migrate from their source in the major groove to reactive sites in the minor groove. Due to the increased lifetime of the reactive species required to access these minor groove sites, abstraction of hydrogen atoms from the bulk solvent becomes a significant competing side reaction (36, 37).

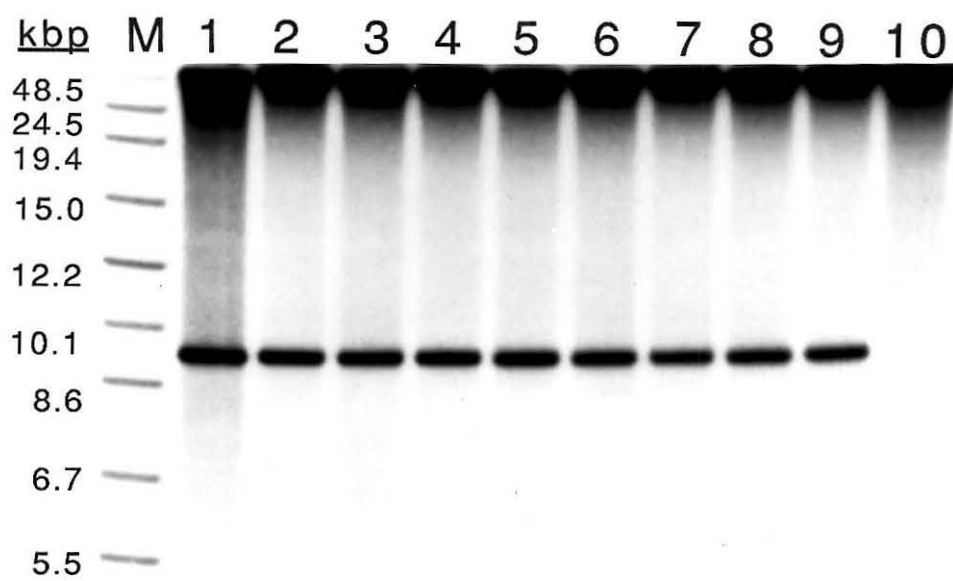
The fundamental limitation to the efficiency of affinity cleaving by oligonucleotide-EDTA•Fe(II) is autocleavage of the oligonucleotide (38, 39). Griffin and Dervan reported significant degradation of T* oligonucleotides in solution during the course of a cleavage reaction when monitored by radiolabeling of the oligomer (38). Ts'o and coworkers also reported significant degradation of oligonucleotides containing EDTA•Fe(II) under

Fig. 2.4. Autoradiogram of a pulse chase competition reaction between oligonucleotides with and without the affinity cleaving agent EDTA•Fe(II). Oligonucleotide-EDTA (1 μ M) was preequilibrated with Fe(NH₄)₂(SO₄)₂•6H₂O (1 μ M) and spermine (1 mM) for five minutes. The metal complexed oligonucleotide was added to R end-labeled bacteriophage λ DNA (4 μ M in base pairs) in 100 mM NaCl, 25 mM tris-HCl pH 7.0, at 24° C. The DNA was incubated with the oligonucleotide for 30 minutes to facilitate triple helix formation. Fifty fold excess (50 μ M) of a non-T* oligonucleotide [5'-(T)₄C(T)₆C-(T)₄CT-3'] identical in sequence to the T* oligomer was added to the triplex solution. DTT (4 mM) was added to the reaction at variable times after the addition of the non-T* oligomer and the cleavage reaction allowed to proceed for six hours. Cleavage products were resolved on a 0.4%, 1xTBE agarose gel. The gel was dried under vacuum and an autoradiogram obtained with an intensifying screen at -70° C.

Lane M, DNA markers of the size indicated. Lane 1, cleavage reaction on bacteriophage λ DNA performed in the absence of competitive non-T* oligonucleotide. DTT was added at t = 0. Lane 2-9, variable times of DTT addition after the addition of non-T* oligonucleotide. Times are t = 0, 10, 20, 30, 45, 60, 90, and 120 min. Lane 10, unreacted control containing all reagents except DTT. The higher cleavage efficiency observed in the absence of non-T* oligonucleotide (Lane 1) suggests that some oligonucleotide exchange might be occurring during the six hour cleavage incubation, or that the high oligonucleotide concentration in lanes 2-9 partially quenches the reaction.

The high concentration of non-T* oligonucleotide in the solution guarantees that if the T* oligomer dissociates from the target site prior to DTT addition, it would not be capable of rebinding. Despite the presence of non-T* oligonucleotide for at least 2 hours, the cleavage reaction still proceeds at high efficiency. This suggests that the oligonucleotide dissociation rate is on the time scale of hours, and that oligonucleotide exchange at the target site is quite slow.

T* Oligo	+	+	+	+	+	+	+	+	+	+
NonT* Oligo	-	+	+	+	+	+	+	+	+	+
DTT (Min)	0	0	10	20	30	45	60	90	120	-



oxidative conditions (39). Autocleavage of an oligonucleotide bound to the DNA duplex would result in a competitive inhibitor of the reaction. The degraded species would retain binding capacity, but would be unable to cleave the target site due to loss of the 5' terminus containing the EDTA•Fe(II) (38).

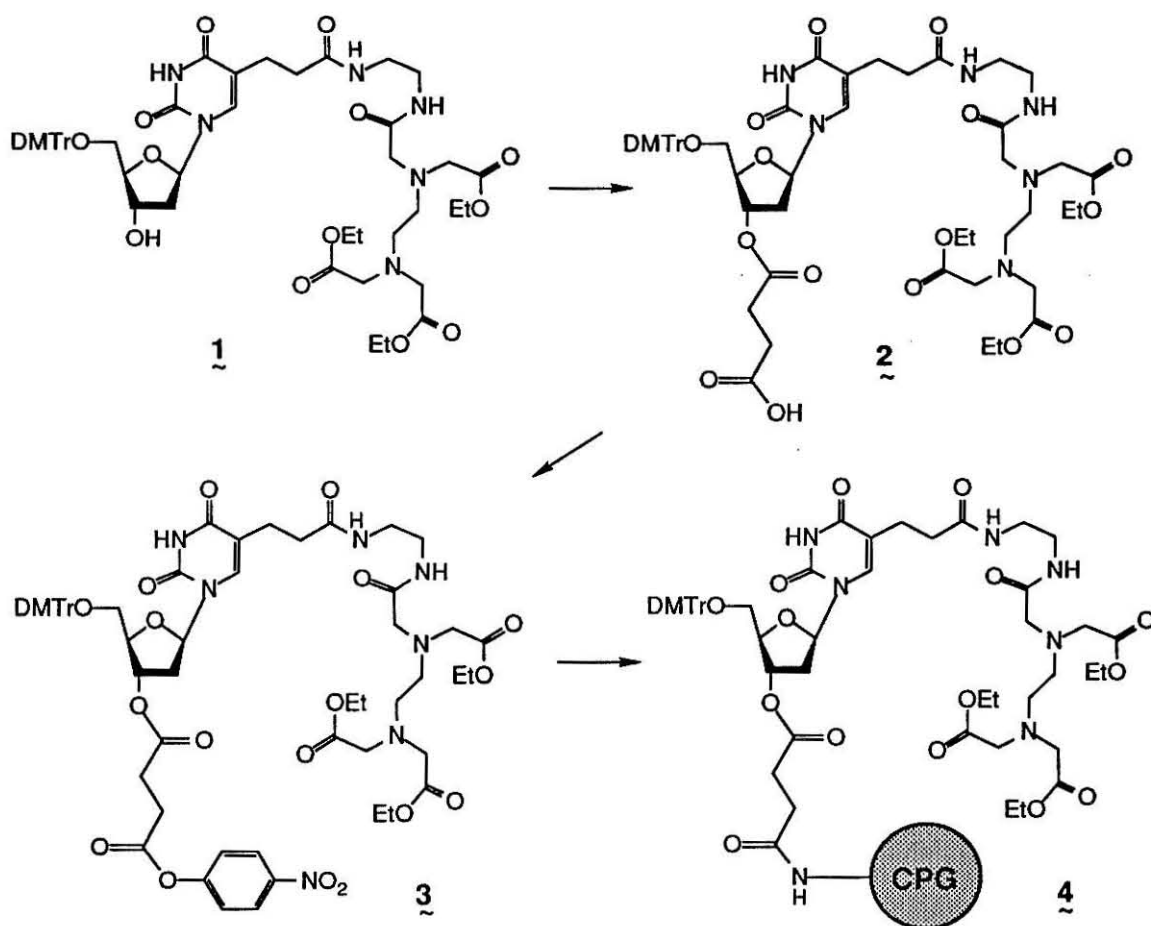


Fig. 2.5. Synthetic scheme for 5'-DMT, 3'-succinyl, thymidine-EDTA-triethyl ester coupled to control pore glass (CPG). See Materials and Methods for detailed protocol.

If autocleavage is the limiting factor for efficient cleavage, one possible strategy to improve the yield is to construct oligonucleotides containing multiple cleaving moieties. This would place several EDTA•Fe(II) groups into a localized region and generate a high concentration of strand nicking at the oligonucleotide binding site.

T* can be introduced at the 5' terminus and at any internal position by coupling T* phosphoramidite to the 5' end of a growing nucleotide chain during automated DNA synthesis (5, 40). Placement at the 3' terminus, however, required preparation of a T* solid support (41). Synthesis of 5'-DMT, 3'-succinyl T* triethylester-control pore glass (CPG) (4) is shown in Fig. 2.5. 5-DMT-T*-triethylester (1) was succinylated at the free 3'-hydroxyl to yield nucleoside 2. Activation of the succinylated T* derivative with p-nitrophenol afforded nucleoside 3. The crude mixture containing the activated T* derivative (3) was added directly to long chain alkyl amine CPG beads to yield 4. Nucleoside loading was determined by DMT deprotection with acid. Free amines on the CPG were capped with acetic anhydride to prevent nonspecific coupling to the support during subsequent oligonucleotide synthesis.

Oligonucleotides (1-6) specific for the homopurine sequence in the bacteriophage λ genome (Fig. 2.6) (4) containing one to four T* moieties at 5', 3' and internal positions were synthesized. The high resolution cleavage pattern of oligonucleotides containing a 3'-T* was examined on a 7.4 kb EcoRI restriction fragment from the bacteriophage λ genome. This fragment contained the homopurine target site 18 bp from the radiolabeled terminus. Oligonucleotide 1 containing a 5'-T* demonstrated a 5' shifted cleavage pattern (Fig. 2.7-1) consistent with major groove binding similar to that reported for a restriction fragment (3). The pattern for the 3' T*

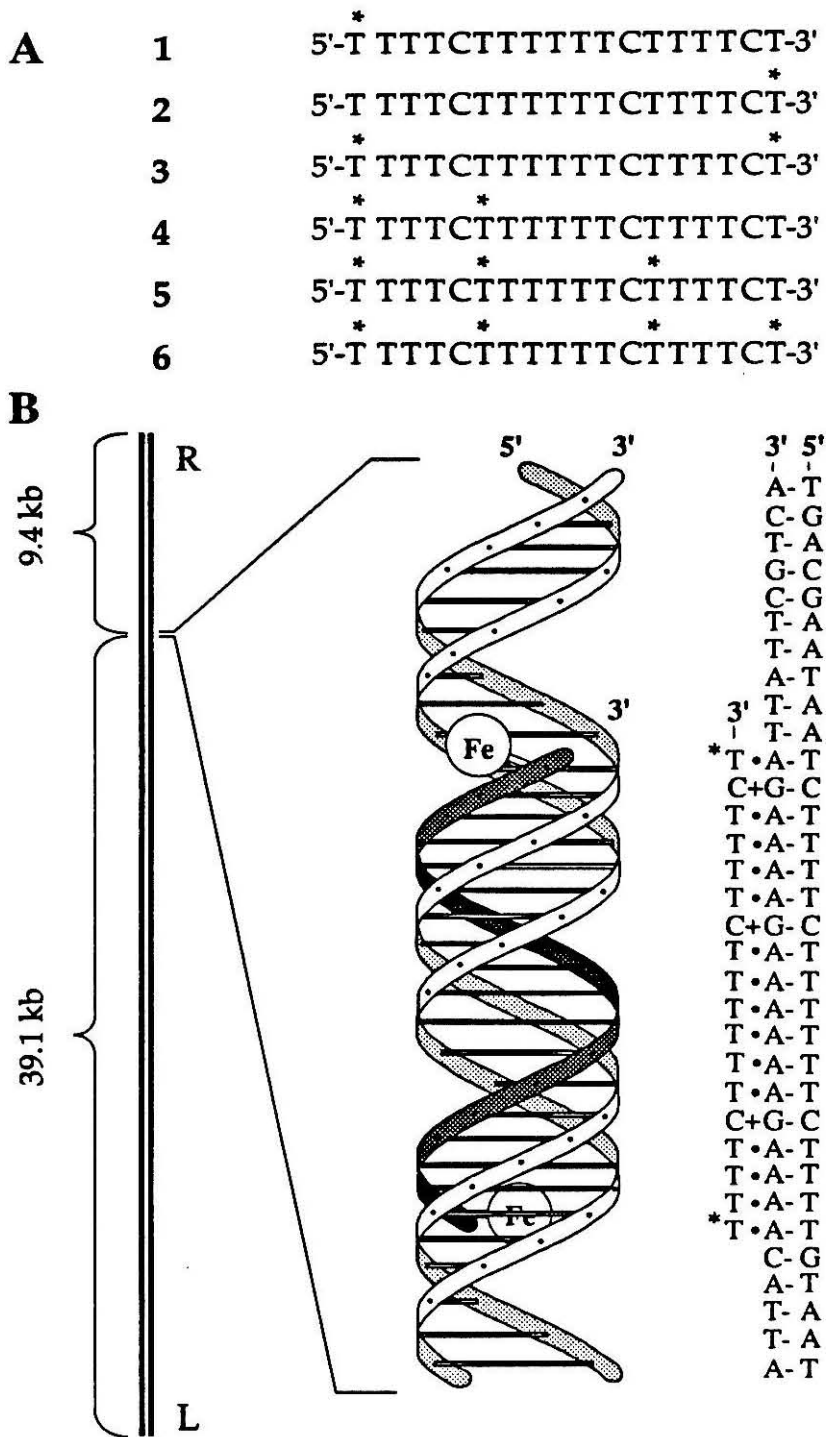


Fig. 2.6. Enhanced cleavage efficiency using oligonucleotides containing multiple T* moieties. **A.** Sequence of oligonucleotides 1-6 specific for triple helix formation at a single homopurine target site in the bacteriophage I genome. **B.** (Left) Double strand cleavage of bacteriophage I DNA generates two fragments, 39.1 (L) and 9.4 kb (R) in size. (Center) Schematic model of triple helix formation between the Watson-Crick homopurine-homopyrimidine site on I DNA and the Hoogsteen hydrogen bonded pyrimidine oligonucleotide 3 with 5' and 3' EDTA•Fe(II).

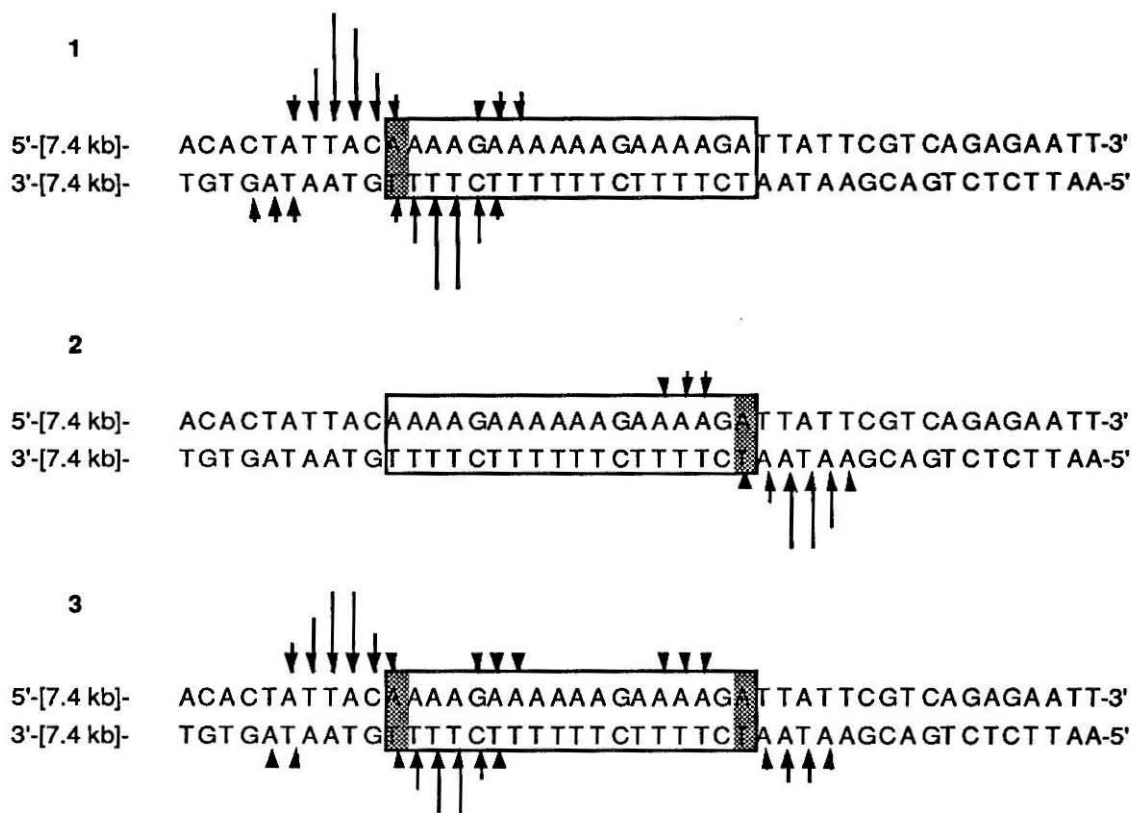


Fig. 2.7. Histogram of the high resolution DNA cleavage pattern on bacteriophage λ DNA by oligonucleotides 1-3 as derived from densitometry of the autoradiogram. The arrow height represents the relative cleavage intensities at the indicated bases. The shaded bases indicate the base bound by T* in each triplex.

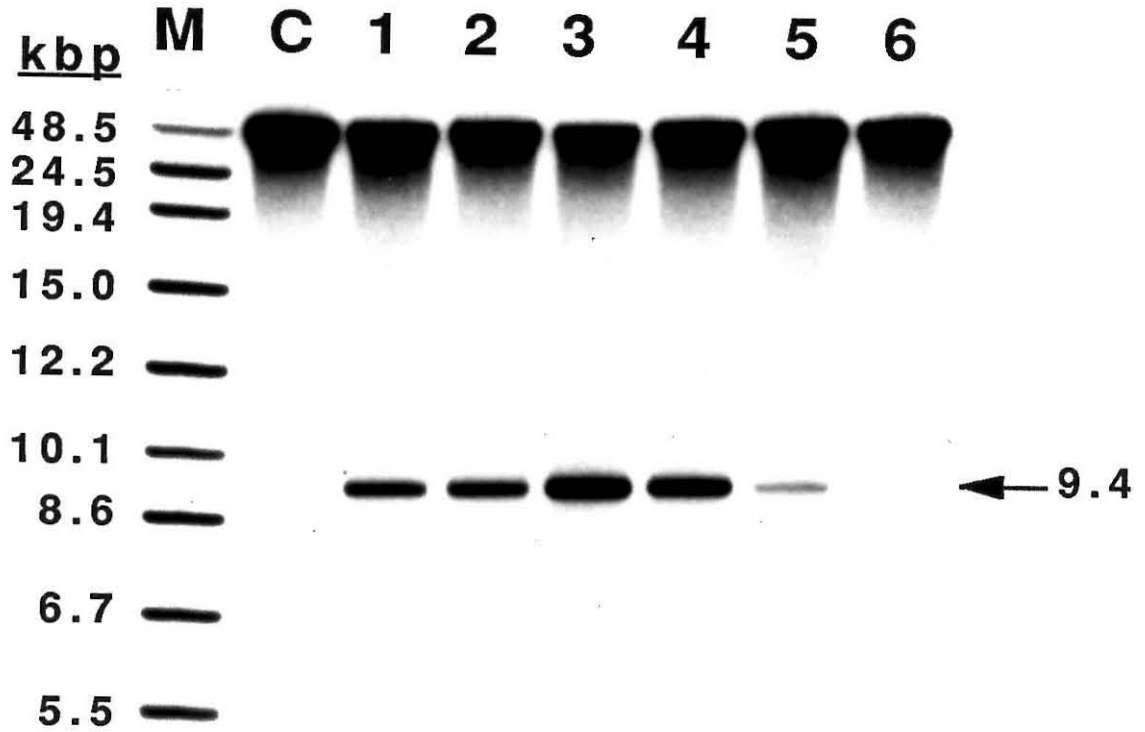
oligonucleotide was also 5' shifted, but the purine strand was cut significantly less efficiently (Fig. 2.7-2). The cleavage pattern of oligonucleotide 3 with T* at both termini, was a composite of the single T* oligonucleotides, although the intensity of cleavage at each base was slightly lower (Fig. 2.7-3).

Differential single strand cleavage efficiency by T* placed at various positions within the Hoogsteen strand indicated structural features of the complex (1). The 5'-T* displayed equal cleavage of the purine and pyrimidine strands. Reduced cleavage of the purine strand by the 3'-T* oligonucleotide 2 is consistent with asymmetric self-footprinting of the DNA. The geometry of the 3'-T* EDTA in the pyrimidine triple helix motif requires that the oxidative species diffuse over the oligonucleotide phosphate backbone to access the purine strand (Fig. 2.6). Thus, the 5' and 3' termini are not geometrically equivalent. Interestingly, Beal and Dervan showed that 3' T* in the purine motif shows a reversed pattern of access to the Watson-Crick strands, i.e., the purine strand is cut efficiently and the pyrimidine strand is footprinted (42). In this motif the third strand is bound in the reverse orientation, closer to the center of the major groove, and employs reverse Hoogsteen TAT triplets (43).

Site specific double strand cleavage efficiencies of the λ genome by oligonucleotides 1-6 were also tested (Fig. 2.8). Oligonucleotide 1 was previously reported to yield $22\pm 3\%$ cleavage (Lane 1) (4). The double strand efficiency of the 3' T* oligonucleotide 2 was slightly lower ($17\pm 3\%$) (Lane 2). Oligonucleotide 3 with T* at both 5' and 3' termini yielded $40\pm 4\%$ cleavage, approximately the sum of cleavage from oligonucleotides 1 and 2 (Lane 3). Oligonucleotide 4 also contained two T* residues but generated only $27\pm 3\%$ cleavage (Lane 4). This may result from lower binding affinity of the probe due to the close proximity of the two T* residues. Oligonucleotides 5 and 6

Fig. 2.8. Site-specific double strand cleavage of bacteriophage λ DNA by oligonucleotides 1-6. The number above each lane indicates the oligonucleotide tested in the reaction. Right end-labeled λ DNA (4 μ M in base pairs) was digested with 1.0 μ M of oligonucleotides 1-6 in one molar equivalent of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ per EDTA, 1 mM spermine, 100 mM NaCl, 25 mM tris-acetate pH 7.0, and 4 mM DTT at 24° C. Oligonucleotide, $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ and spermine were preequilibrated together for 5 minutes, added to the DNA in buffered salt solution, and incubated an additional 30 minutes to facilitate triple helix formation. DTT was added to a final concentration of 4 mM, and the cleavage reaction was allowed to proceed for 6 hours. Cleavage products were resolved on a 0.4%, 1xTBE agarose gel. The gel was dried under vacuum and an autoradiogram obtained with an intensifying screen at -70° C.

Lane M, DNA size markers of size indicated. Lane C, intact right end-labeled λ DNA control. Lanes 1-6, cleavage reactions using oligonucleotide 1-6, respectively. Arrow on the right indicates the 9.4 kb cleavage fragment.



(with three and four T* residues, respectively) failed to cut efficiently (<5%) (Lanes 5 and 6). Cleavage by oligonucleotides 5 and 6 was enhanced under more permissive binding conditions, but the effect was small, and cleavage remained lower than that obtained with either single or double T* oligonucleotides. This suggests that the oligonucleotides did not bind the target site, and indicates that binding affinity was reduced by multiple T* substituents. Decreased cleavage may reflect triplex destabilization due to unfavorable steric interactions in the major groove involving bulky EDTA substitutions of thymidine at C5. It further suggests that a critical number of base triplets were necessary to stabilize each T* residue in the probe (38).

Mathematical Model of Autocleavage. In spite of oligonucleotide saturation of the target site and the introduction of multiple cleaving functionalities into a single oligonucleotide, the efficiency of affinity cleaving remains fundamentally limited by oligonucleotide self cleavage. Statistically, double strand cleavage efficiency is limited by the probability of cleaving both duplex DNA strands before suicide inactivation of the probe. A simple mathematical model expressing the efficiency of double strand cleavage (E) is the summation:

$$E = \sum_{i=2}^{\infty} py^{(i-1)} \cdot pu + pu^{(i-1)} \cdot py \quad (\text{Eq. 2.1})$$

where detectable oxidative events, i , are defined as those resulting in cleavage of the purine, pyrimidine, or Hoogsteen strands. The corresponding probabilities per event are pu , py , and H , respectively. For each detectable oxidative event, $pu + py + H = 1$. An analysis using this model was applied to oligonucleotide 1 containing the 5' T*. For this oligonucleotide, the observed pu/py ratio was 1.0 (Fig. 2.4), and the double strand cleavage efficiency (E) was 25%. Iteration using the above equation to solve for the cleavage probabilities

yielded values of 0.29, 0.29, and 0.42 for pu , py , and H , respectively. This suggests that for each oxidative event, autocleavage is 1.4 times more probable than cleavage of either the purine or pyrimidine strands.

In the case of oligonucleotide **2** with a 3'-T*, the pu/py ratio was reduced to 0.3, while the double strand cleavage efficiency was 17%. The corresponding values for pu , py , and H were 0.13, 0.46, and 0.41, respectively. Thus, pyrimidine and Hoogsteen strands were cleaved with approximately equal probability, while the purine strand was cut less efficiently due to protection by the bound strand. Partial protection of the purine strand results in a lower probability of double strand cleavage before the oligonucleotide was inactivated. Therefore, the difference between the values obtained for oligonucleotides **1** and **2** reflect the distinct steric environments at the 5' and 3' termini. This simple model allows quantitative estimation of the significance of autocleavage as a limitation to double strand cleavage efficiency.

In addition to providing a simple mechanism to improve the efficiency of double strand cleavage, the coupling of T* to the CPG solid support offers a number of synthetic advantages. T* phosphoramidite coupling to the nascent nucleotide chain typically proceeds with only moderate coupling efficiencies (50-75%), because the reaction is extremely sensitive to the trace impurities typically present in the phosphoramidite. Such impurities significantly reduce shelf life of the T* phosphoramidite making it necessary to freshly prepare the material for each synthesis. Coupling to the 5' end of the nascent oligonucleotide chain is also inherently wasteful due to the need for 20-fold excess nucleoside for efficient coupling. Conversely, the T*-CPG consistently couples to high efficiency, does not require phosphitylation for activation, and can produce ten fold more couplings per gram than in the T*

phosphoramidite form. T*-CPG support is stable at 4° C for at least two years, and 1.0 g of protected T* coupled to T*-CPG support is sufficient to make several hundred oligonucleotides.

Enhanced cleavage of bacteriophage λ DNA by pH cycling. Despite the improved cleavage efficiency produced by chemical modification of the oligonucleotide, binding affinity and autocleavage still limit the overall cleavage efficiency. A second method to improve the cleavage yield was by performing multiple rounds of cleavage. Because the majority of the binding site remained occupied despite autocleavage of the oligonucleotide, and the inactivated autocleavage product of autocleavage could readily bind to the target site after denaturation, it was necessary to physically remove the third strand from the substrate DNA between each round of reagent addition. Several attempts to use increased temperature to remove the third strand proved unsuccessful as the elevated temperature promoted strand nicking by oxidative reagents in the solution, and no overall cleavage enhancement was observed at the target site. Ethanol coprecipitated the 48.5 kb substrate and the 18 base oligomer. Triple helix formation in the pyrimidine motif is highly pH sensitive, and providing a nondestructive mechanism for triple helix disruption (3, 4, 24, 44-47). Separation could be achieved by embedding the DNA substrate in agarose and diffusing the oligomer into and out of the matrix as a function of pH (4). The size differential between the oligonucleotide and genomic substrate would facilitate efficient separation.

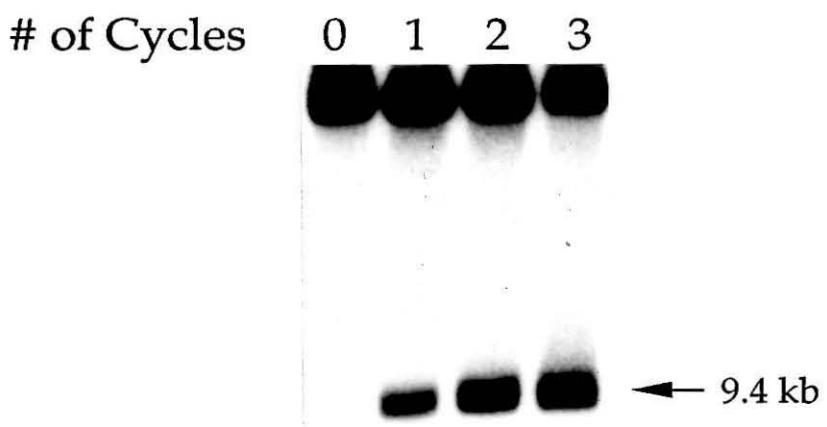
Three cleavage cycles were performed on bacteriophage λ DNA using high pH washes between cleavage cycles to remove the oligonucleotide. Increased efficiency was observed after each round of cleavage (Fig. 2.9) Although DNA embedded in agarose was generally cut slightly less efficiently than in solution, pH cycling resulted in improved cleavage yields. Using

oligonucleotide 1, the efficiency improved from 18% after one pH cycle, to 29% after two, and 37% cleavage after three cycles (Fig. 2.9). The yields using oligonucleotide 2 also improved with cycling, but they were not significantly better than observed in solution with a single round of cleavage. The values were 19% after one cycle, and 37% after two cycles.

Although theoretically, the cleavage rounds could be continued until the substrate DNA was cut quantitatively, incomplete removal of reagents from the previous cycles resulted in gradually lower efficiency in later cleavage rounds. Additionally, the nonspecific oxidative nicking became significant after the first two cycles. Despite the preference of performing the reactions in solution for both ease and reduced levels of random DNA degradation, pH cycling was a useful method for improving the efficiency of cutting large chromosomal DNA that must be manipulated in agarose (see Chapter 3) (48).

Fig. 2.9. Improved cleavage efficiency of bacteriophage λ DNA by pH cycling. P^{32} right end-labeled λ DNA embedded in 0.8% low melting point agarose ($\approx 100 \mu\text{L}$ total plug volume) was equilibrated in 1.0 ml of triplex buffer (NTS: 100 mM NaCl, 25 mM tris-HCl pH 7.0, and 1.0 mM spermine \cdot HCl) for 15 minutes, decanted and overlaid with 120 μL of NTS. 2 μL of 100 μM oligonucleotide 1 was preequilibrated with an equal volume of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ and 3 μL of the resulting solution added to the plug. The mixture was incubated for 3.0 hours at 24 $^\circ$ C to facilitate diffusion of the oligonucleotide into the agarose and to promote triple helix formation. The NTS was then decanted and a fresh 120 μL aliquot of NTS added to the plug. 3 μL of 0.2 M DTT was added and the reaction incubated an additional 7 hours. The reaction was stopped by the addition of 1.0 ml of 100 mM tris-HCl pH 8.5 to disrupt the triplex. The oligonucleotide was removed from the solution by three 1.0 ml washes (15 min. to overnight) with 100 mM tris-HCl pH 8.5. The plugs were then reequilibrated in NTS buffer and the reaction repeated as described. After 1, 2, or 3 rounds of cleavage were complete, the plugs were equilibrated in 10 mM tris-HCl, 1 mM EDTA pH 8.0 and loaded onto a 0.5% agarose gel for resolution of the cleavage products. The gel was dried under vacuum and an autoradiogram obtained with an intensifying screen overnight at -70 $^\circ$ C.

Lane C, No treatment control of intact right end-labeled λ DNA. Lanes 1-3, cleavage reactions using oligonucleotide 1 with 1, 2 and 3 pH cycles, respectively. The 9.4 kb cleavage product is indicated with an arrow to the right.



Part II

Cooperative Site Specific Binding of Oligonucleotides to Duplex DNA

Cooperative binding to DNA has been observed in many natural systems (for reviews see references 49-51). Cooperative interactions between DNA binding proteins result in greater sequence specificity and improved sensitivity to changes in concentration. The activity of a number of DNA binding proteins is affected by cooperative interactions with homologous or heterologous proteins bound at neighboring recognition sites (52-60). The cooperative interaction is typically through protein-protein contacts between domains distinct from those for DNA binding (52-60).

Cooperativity has also been implicated in binding of small molecules to DNA (49). The interaction is either between the DNA binding molecules or through a cooperative allosteric change in the DNA (49). Examples of the first are the heterodimeric binding of distamycin and 2-imidazole-netropsin in the minor groove of DNA (61), and the cooperative intercalation of acridine (62). In the second cooperative model, the molecules interact indirectly by altering the conformational state of the nucleic acid. Examples of this include ethidium (63), daunomycin (64), actinomine (65), or actinomycin D (65) cooperatively binding to left-handed Z DNA producing a local conformational change to a right-handed helix.

Cooperative binding of single stranded oligonucleotides to single stranded templates has also been examined. Studies on the binding of oligoinosine I(pI)₅₋₁₀ to a polycytidine template found that the oligomer cooperatively bound in clusters (66). Cooperativity was also detected for the

binding of sequence specific oligonucleotides to adjacent target sites on single stranded RNA or DNA templates (39, 67, 68).

To define the degree of cooperative interaction between two heterologous ligands, it is necessary to determine binding isotherms for one ligand in the presence and in the absence of the second ligand (56). A qualitative estimate can be made, however, by observing differential cleavage efficiencies by affinity cleaving. This analysis was used to examine cooperative interactions between heterologous oligonucleotides bound to adjacent sites on double stranded DNA by triple helix formation (7).

Reprinted from the Journal of the American Chemical Society, 1989, 111, 7286.
Copyright © 1989 by the American Chemical Society and reprinted by permission of the copyright owner.

Cooperative Site Specific Binding of Oligonucleotides to Duplex DNA

Scott A. Strobel[†] and Peter B. Dervan*

Arnold and Mabel Beckman Laboratories of Chemical
Synthesis, California Institute of Technology
Pasadena, California 91125

Received March 20, 1989

Cooperative interactions between DNA binding ligands are critical to their specificity, affinity, and biological activity.¹⁻⁴ Triple helix formation by oligonucleotides is the most powerful chemical approach to date for the sequence-specific recognition of double helical DNA.⁵⁻⁹ Hoogsteen hydrogen bonded base triplets, TAT and C+GC, result from pyrimidine oligonucleotides binding site specifically to purine duplex sequences. In the triple helical model, a binding site size of 18 purine base pairs affords 36 discrete sequence-specific hydrogen bonds for recognition of DNA in the major groove. As a possible mechanism for improving the specificity of triple helix formation, we tested whether oligonucleotides could cooperatively bind to a double-stranded DNA template.

We report that two different pyrimidine oligonucleotides, which are nine bases in length, cooperatively bind to an 18 base-pair homopurine site in bacteriophage λ genomic DNA by triple helix formation. The purine target sequence 5'-A₄GA₂GA₄GA-3' occurs once in λ DNA¹⁰ (48.5 kilobase pairs) and can be considered as two contiguous unique half-sites, 5'-A₄GA₄-3' and 5'-A₂GA₄GA-3'.

[†] Howard Hughes Medical Institute Doctoral Fellow.

(1) For cooperative binding of homologous and heterologous proteins on DNA, see: (a) Johnson, A. D.; Meyer, B. J.; Ptashne, M. *Proc. Natl. Acad. Sci. U.S.A.* 1979, 76, 5061. (b) Minter, S. J.; Clore, G. M.; Gronenborn, A. M.; Davies, R. W. *Eur. J. Biochem.* 1986, 161, 727. (c) Schule, R.; Muller, M.; Murakami, H. O.; Renkawitz, R. *Nature* 1988, 332, 87. (d) Giniger, E.; Ptashne, M. *Proc. Natl. Acad. Sci. U.S.A.* 1988, 85, 382. (e) Poellinger, L.; Yoza, B. K.; Roeder, R. G. *Nature* 1989, 337, 573. (f) Ren, Y. L.; Gargus, S.; Adhya, S.; Krakow, J. S. *Proc. Natl. Acad. Sci. U.S.A.* 1988, 85, 4138. (g) Hirano, A.; Wong, T. *Mol. Cell. Biol.* 1988, 8, 5232. (h) For reviews, see: Ptashne, M. *Nature* 1988, 335, 683.

(2) For cooperative binding of small molecules on DNA, see: (a) Hogan, M.; Dattagupta, N.; Crothers, D. M. *Nature* 1979, 278, 521. (b) Dattagupta, N.; Hogan, M.; Crothers, D. M. *Biochemistry* 1980, 19, 5998. (c) Graves, D. E.; Krugh, T. R. *Biochemistry* 1983, 22, 3941. (d) Walker, G. T.; Stone, M. P.; Krugh, T. R. *Biochemistry* 1985, 24, 7462. (e) Walker, G. T.; Stone, M. P.; Krugh, T. R. *Biochemistry* 1985, 24, 7471. (f) Chaires, J. B. *Biochemistry* 1985, 24, 7479. (g) Rosenberg, L. S.; Carvlin, M. J.; Krugh, T. R. *Biochemistry* 1986, 25, 1002. (h) Hardin, C. C.; Walker, G. T.; Tinoco, I. *Biochemistry* 1988, 27, 4178. (i) For a review, see: Wilson, D. W. *Progress Drug Res.* 1987, 31, 193.

(3) For cooperative binding of oligonucleotides to single-stranded DNA templates, see: (a) Tazawa, I.; Tazawa, S.; Ts'o, P. O. P. *J. Mol. Biol.* 1972, 66, 115. (b) Springgate, M. W.; Poland, D. *Biopolymers* 1973, 12, 2241. (c) Asseline, U.; Delarue, M.; Lancelot, G.; Toulme, F.; Thuong, N. T.; Garestier, T. M.; Helene, C. *Proc. Natl. Acad. Sci. U.S.A.* 1984, 81, 3297. (d) Lin, S. B.; Blake, K. R.; Miller, P. S.; Ts'o, P. O. P. *Biochemistry* 1989, 28, 1054.

(4) For cooperative binding of sequence-specific oligonucleotides to single-stranded RNA templates, see: (a) Maher, L. J.; Dolnick, B. J. *Arch. Biochem. Biophys.* 1987, 253, 214. (b) Maher, L. J.; Dolnick, B. J. *Nucl. Acids Res.* 1988, 16, 3341.

(5) (a) Moser, H. E.; Dervan, P. B. *Science (Washington, D.C.)* 1987, 238, 645. (b) Strobel, S. A.; Moser, H. E.; Dervan, P. B. *J. Am. Chem. Soc.* 1988, 110, 7927. (c) Fovsic, T.; Dervan, P. B. *J. Am. Chem. Soc.* 1989, 111, 3059. (d) Maher III, L. J.; Wold, B.; Dervan, P. B. *Science (Washington, D.C.)* 1989, in press. (e) Griffin, L. C.; Dervan, P. B. *Science* 1989, in press.

(6) (a) Doan, T. L.; Perrouault, L.; Praseuth, D.; Habboub, N.; Decout, J. L.; Thuong, N. T.; Lhomme, J.; Helene, C. *Nucl. Acids Res.* 1987, 15, 7749. (b) Praseuth, D.; Perrouault, L.; Doan, T. L.; Chassignol, M.; Thuong, N.; Helene, C. *Proc. Natl. Acad. Sci. U.S.A.* 1988, 85, 1349.

(7) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* 1988, 241, 456.

(8) Rajagopal, P.; Feigon, J. *Nature* 1989, 339, 637.

(9) For a recent review on triple helical polynucleotide structures, see: Wells, R. D.; Collier, D. A.; Hanvey, J. C.; Shimizu, M.; Wohlrab, F. *FASEB J.* 1988, 2, 2939.

(10) Sanger, F.; Coulson, A. R.; Hong, G. F.; Hill, D. F.; Petersen, G. B. *J. Mol. Biol.* 1982, 162, 729.

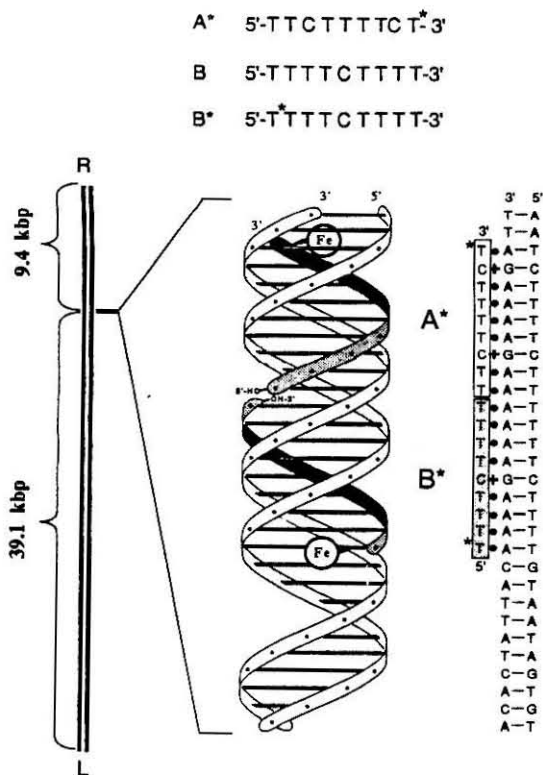


Figure 1. Upper: pyrimidine oligonucleotides designed to bind adjacent half-sites of the target sequence. A* and B* contain an EDTA cleaving function at their 3' and 5' termini, respectively. Lower: double strand cleavage of λ DNA affords two fragments, 39.1 and 9.4 kbp in size. Simplified models of the triple helix complex between adjacent Watson-Crick homopurine-homopyrimidine half-sites and the Hoogsteen pyrimidine oligonucleotides A* and B* bound head-to-tail.

Two pyrimidine oligonucleotides, specific for the adjacent half sites, were tested for cooperative binding interaction by the affinity cleaving method.^{5,11,12}

Oligonucleotides A*, B, and B* were synthesized by automated methods (Figure 1). A modified nucleoside for cleaving DNA, thymidine-EDTA (T*),¹² was placed at the 3' end of 5'-T₂CT₄CT*-3', (A*).¹³ Oligonucleotides 5'-T₄CT₄-3' were synthesized with and without a T* moiety at the 5' end (B* and B, respectively). λ DNA¹² was labeled with ³²P at the right end (R) with AMV reverse transcriptase.^{5b} Specific double strand oxidative cleavage at the target site by an oligonucleotide-EDTA-Fe affords a 9.4 kbp fragment (Figure 2).^{5b} The intensity of the 9.4 kbp fragment indicates the extent of oligonucleotide binding.

Combinations of oligonucleotides A*, B, and B* at 5 μ M concentrations were incubated at 24 °C with Fe(NH₄)₂(SO₄)₂, spermine (1.0 mM), λ DNA (approximately 4 μ M in base pairs), 100 mM NaCl, and 25 mM tris-acetate in 17% yield under previously described conditions.^{5b} Strobel, S. A.; Dervan, P. B., unpublished work.

(11) Dervan, P. B. *Science* 1986, 232, 464.

(12) Dreyer, G. B.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* 1985, 82, 968.

(13) Thymidine-EDTA (T*)¹² was placed at the 3' terminus of oligonucleotide A* by coupling 5' DMT-thymidine EDTA-triethyl ester to a control pore glass support. Eighteen base oligonucleotides with T* at the 3' end are capable of generating double strand cleavage in 17% yield under previously described conditions.^{5b} Strobel, S. A.; Dervan, P. B., unpublished work.

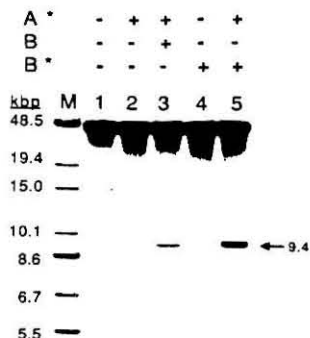


Figure 2. Site-specific double strand cleavage of bacteriophage λ DNA. Autoradiogram of ³²P right end-labeled λ DNA, 48.5 kbp in size, on an 0.5% agarose gel. Lane M, DNA size markers obtained by restriction enzyme digestion of left end-labeled DNA with BamHI, Apa I, and Sma I; digestion of right end-labeled DNA with BamHI, Sma I, and Xho I; 48.5 kbp (undigested DNA), 19.4, 15.0, 10.1, 8.6, 6.7, and 5.5 kbp. Lane 1, right end-labeled intact λ DNA control. Lane 2, right end-labeled λ DNA with 5.0 μ M A* and 5 μ M Fe(II). Lane 3, λ DNA with 5.0 μ M A*, 5.0 μ M B, and 5.0 μ M Fe(II). Lane 4, λ DNA with 5.0 μ M B* and 5.0 μ M Fe(II). Lane 5, λ DNA with 5.0 μ M A*, 5.0 μ M B*, and 10 μ M Fe(II). Arrow on the right indicates 9.4 kbp cleavage fragment.

reactions proceeded for 6 h (24 °C) and were stopped by ethanol precipitation. Double strand cleavage products were separated by agarose gel electrophoresis and visualized by autoradiography (Figure 2). Cleavage efficiency was quantitated by scintillation counting of the individual bands.

A* incubated in the absence of B generated a marginally detectable level of cleavage, 1.3 (\pm 0.2)% (Figure 2, lane 2). A similar result 1.7 (\pm 0.2)% was observed for B* incubated alone, (lane 4). This suggests that these nine base pyrimidine oligonucleotides have low binding affinity for their target sites under the conditions chosen (24 °C, pH 6.6), consistent with previous observations.^{5b} In contrast, when A* was incubated in the presence of B, the cleavage efficiency improved to 4.4 (\pm 0.5)% (lane 3). This 3.5-fold improvement in double strand cleavage indicates a positive binding interaction between contiguous oligonucleotides A* and B, aligned head-to-tail in the major groove of duplex DNA. Double strand cleavage can be further increased to 10.5 (\pm 1.1)% when both oligonucleotides were equipped with EDTA (A* in the presence of B*, lane 5).

Cooperative binding could arise from two different interactions: base stacking of the terminal thymine bases of the Hoogsteen strand in the major groove and/or an induced conformational change at contiguous DNA sites. There is literature precedent for both types of interactions, although in significantly different systems.²⁻⁴ If base stacking between the terminal bases of contiguous oligonucleotides in the triple helical complex is important for cooperative binding,¹³ then disruption of this interaction could be utilized to make sequence-specific recognition highly sensitive to single base mismatches. Design of structural motifs that enhance the cooperative interaction of adjacent sequence-specific binding oligonucleotides on duplex DNA is in progress.

Acknowledgment. This work was supported in part by the National Institutes of Health (GM-35724) and the Caltech Consortium in Chemistry and Chemical Engineering (Founding members: E. I. du Pont de Nemours & Co., Inc., Eastman Kodak Company, Minnesota Mining and Manufacturing Company, and Shell Oil Company Foundation).

(14) Strain c1857 *ind1 Sam7* from New England Biolabs.

(15) (a) Solie, T. N.; Schellman, J. A. *J. Mol. Biol.* 1968, 33, 61. (b) Haran, T. E.; Berkovich-Yellin, Z.; Shakked, Z. *J. Biomol. Struct. Dyn.* 1984, 2, 397.

Nature of the Cooperative Interaction. We originally proposed two possible mechanisms to explain the cooperative effect observed with these oligonucleotides: (i) base stacking of the terminal thymine bases of the Hoogsteen strand in the major groove, and/or (ii) an induced allosteric change in the DNA at contiguous binding sites (7). While we can not rule out either of these explanations, recent data demonstrates the importance of base stacking for cooperativity. Horne and Dervan incorporated an abasic deoxyribose residue into a full length oligonucleotide and observed a significant reduction in binding affinity (69). Distefano and Dervan did not observe cooperative binding if the ends of the oligonucleotides were separated by one base (70). Povsic and Dervan failed to detect end-to-end cooperativity if the oligonucleotides were abutting, but on opposite faces of the major groove (33). This configuration would eliminate base stacking while retaining contiguous binding. In a similar study, Beal and Dervan failed to detect cooperativity between two oligonucleotides that bound adjacent sites in the pyrimidine and purine motifs, respectively (42). This configuration also minimized base stacking while retaining contiguous binding. Together these data suggest that cooperativity is primarily through base stacking between the terminal bases, or possibly hydrogen bonding between the terminal deoxyribose sugars (71, 72).

In 1989 we reported a 3.5 fold cooperative enhancement. Subsequent to this analysis cooperative binding isotherms were studied by a restriction enzyme assay (34) in a system more compatible with analysis of cooperativity. Ratios as high as ten to one were obtained at optimal concentrations on the binding isotherm (73). Povsic and Dervan also assayed end-to-end cooperativity by alkylation and observed similar values (33).

Artificial Motifs for Cooperative Triple Helix Formation. Other cooperative systems involving oligonucleotide directed triple helix formation have now been demonstrated. Povsic and Dervan observed cooperative binding to yeast genomic DNA (33). Distefano and Dervan observed greater than a 40-fold increase in binding affinity due to cooperativity between two oligonucleotides capable of forming a Watson-Crick duplex dimerization domain perpendicular to the triple helix (70, 74). Gates and Dervan observed a modest cooperative effect of 3.7 using a metalloregulated domain for dimerization (75). In this system, oligonucleotides equipped with 1, 10 phenanthroline at their 5' terminus bound cooperatively to a duplex DNA template only upon the addition of cobalt, cadmium or zinc.

Conclusion. Bacteriophage λ genomic DNA served as a valuable substrate for the study of triple helix formation in the critical size range between plasmid DNA and the complex genomes of yeast and human (Chapters 3-4). Cleavage efficiencies as high as 45% were detected at a single target site with low background. Efficient cleavage of DNA embedded in agarose (37%) was also demonstrated. The stringency of triple helix formation was examined as a function of pH, temperature, and oligonucleotide base composition. Cooperative binding of oligonucleotides to duplex DNA via triple helix formation was also demonstrated.

Material and Methods

^1H nuclear magnetic resonance (NMR) spectra were recorded on a Jeol JNM-GX400 FT 400 MHz spectrometer and are reported in parts per million (ppm) from tetramethylsilane. Ultraviolet-visible (UV-Vis) spectra were recorded on a Perkin-Elmer Lambda 4C UV/Vis spectrophotometer. Flash chromatography was carried out under positive air pressure using EM Science Kieselgel 60 (230-400 Mesh). Thin-layer chromatography (TLC) was performed with precoated 0.25 mm silica gel 60 F-254 TLC plates (EM Reagents, Darmstadt F. R. G.). Long chain alkyl amine control pore glass beads (CPG) were obtained from Sigma. T* and T* diisopropyl amino β -cyanoethyl phosphoramidites were synthesized as described (5). Oligonucleotide synthesis was performed on a Beckman System 1 Plus DNA Synthesizer using standard β -cyanoethyl phosphoramidite chemistry and reagents purchased from Beckman (41). 5-methylcytosine and 5-bromouracil β -cyanoethyl phosphoramidites were purchased from Cruachem. Aqueous 5'-[α - ^{32}P]dATP, 5'-[α - ^{32}P]dGTP and 5'-[γ - ^{32}P]dATP (3000 Ci/mmol; 1 Ci = 37 GBq) were purchased from Amersham. Radioactivity was measured with a Beckman LS 3801 liquid scintillation counter. Unlabeled deoxynucleotriphosphates were purchased from Boehringer Mannheim. Bacteriophage λ DNA (cI857ind1Sam7) and all enzymes were obtained from New England Biolabs except AMV reverse transcriptase and calf intestinal alkaline phosphatase, which were purchased from United States Biolabs and Boehringer Mannheim, respectively. Solutions of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$, and dithiothreitol (DTT) were freshly prepared.

Synthesis of 5'-DMT, 3'-succinyl T triethyl ester control pore glass:*

Nucleoside 2. Succinic anhydride (150 mg, 1.5 mmol) was slowly added to a solution of 5'-DMT-T* triethylester (**1**) (1.007 g, 1.0 mmol) and 4-dimethylaminopyridine (DMAP) (60 mg, 0.5 mmol) in dry pyridine (4 mls) and stirred under argon for six hours. A second portion of DMAP (60 mg, 0.5 mmol) and succinic anhydride (100 mg, 1.0 mmol) was added and complete conversion to nucleoside **2** was obtained after 12 hours. The solution was reduced to a gum by rotary evaporation and the residual solvent removed by coevaporation with dry toluene. The crude mixture was dissolved in CH₂Cl₂ (5 mls), washed with 10% cold sodium citrate (2x 5 mls) and H₂O (2x5 mls), dried over anhydrous Na₂SO₄, and reconcentrated. The residue was redissolved in CH₂Cl₂ (5 mls), precipitated in stirring hexane (250 mls), decanted, and dried under vacuum to give nucleoside **2** (1.023 g, 0.93 mmol, 92% yield). ¹H NMR (CDCl₃): δ 9.67 (1H, bs), 8.46 (1H, m), 7.48 (1H, s, H₆), 7.39 (2H, d), 7.32-7.22 (7H), 6.88 (1H, m), 6.84 (4H, d), 6.24 (1H, t, H₁), 5.35 (1H, m, H₃), 4.15 (6H, m, OCH₂CH₃), 4.11 (1H, m, H₄), 3.78 (6H, s, OCH₃), 3.53 (4H, s), 3.41-3.27 (8H), 2.96 (2H, s), 2.78 (4H, m), 2.76 (4H, m), 2.31-2.18 (7H), 1.24 (9H, t, OCH₂CH₃). TLC (5% MeOH in CH₂Cl₂): R_f=0.10.

Nucleoside 4. A solution of nucleoside **2** (1.023 g, 0.93 mmol) in anhydrous dioxane (4 mls) and pyridine (0.2 mls) was stirred with p-nitrophenol (140 mg, 1 mmol) and dicyclohexylcarbodiimide (515 mg, 2.5 mmol) for 2 hours to yield nucleoside **3**. (R_f=0.75 in 10% MeOH, CHCl₃). The crude mixture was filtered through a fritted funnel, washed with dimethylformamide (DMF) (5 mls) and added to CPG beads (5.0 g). Triethylamine was added and the slurry was swirled periodically for 8 hours. The solution was decanted, an additional portion of CPG added (2.5 g), and the coupling reaction allowed to proceed for 12 hours. The derivatized CPG

was washed successively with DMF, followed by MeOH, and finally with ether. The T*-CPG was dried under vacuum. Nucleotide loading was determined to be 29 μ moles of nucleoside per gram of support by a DMT deprotection assay (41). The uncoupled amine groups on the CPG support (7.5 g) were capped in pyridine (30 mls) with acetic anhydride (2 mls) and DMAP (100 mg, 0.8 mmol) for 30 minutes. The support was then refiltered, washed, and dried as above. T* derivatized CPG (25 mg) was loaded into 1.0 μ mole fritted columns for automated oligonucleotide synthesis.

Oligonucleotide Synthesis. Oligonucleotides were synthesized by standard automated methods using β -cyanoethyl phosphoramidites. Oligonucleotides with a 3' T* were synthesized using T* derivatized CPG support. T* nucleotides at other locations were introduced as activated β -cyanoethyl phosphoramidites as described (5). Oligonucleotides were deprotected with 0.1 N NaOH (1.5 mls) for 24 hours at 55° C. The oligomers were purified on a 20% denaturing polyacrylamide slab gel (19:1 bisacrylamide crosslinked) containing 45% urea. The products were visualized by UV shadowing, excised from the gel, extracted with 100 mM NaCl, 1 mM EDTA for 40 hours, and dialyzed against distilled water at 4°C for 3-4 days. The water was changed every 6-10 hours. Oligonucleotide concentration was determined by UV absorption at 260 nm. The extinction coefficient for each oligonucleotide was calculated using the following nucleotide extinction coefficients: T, T*=9000 M⁻¹, C=8000 M⁻¹, BrU=4600 M⁻¹, and MeC=5700 M⁻¹. The MeC^{BrU} oligonucleotides were slightly shifted towards the visible with a λ_{MAX} of 277 nm compared to a λ_{MAX} for the CT oligonucleotide of 267 nm.

Preparation of End Labeled DNA for High Resolution Analysis.

Bacteriophage λ DNA was digested to completion with EcoRI to produce 21.2, 7.4, 5.8, 5.6, 4.9, and 3.5 kb products. The ends of all fragments were 5' or 3'

end labeled. The digest was 3'-end labeled with [α - 32 P]dATP by Klenow fragment of DNA polymerase. 5'-End labeling was performed by alkaline phosphatase treatment to remove the 5' terminal phosphates, followed by phosphoryl transfer with [γ - 32 P]dATP and polynucleotide kinase.

Unincorporated mononucleotides were removed by G-50 Sephadex spin column chromatography in water. The 3' or 5' end labeled 7.4 kb fragment containing the homopurine target site was isolated on a 0.5% agarose gel, extracted by electroelution, ethanol precipitated, and dissolved in 250 mM NaCl and 62 mM tris-acetate pH 7.0.

Preparation of End Labeled DNA for Low Resolution Analysis.

Bacteriophage λ DNA was uniquely labeled at the right or left terminus by filling in the single stranded cos ends. This was done by providing the polymerase with only the appropriate mononucleotides for partially filling in one end. In the absence of all four nucleotides, many polymerases have extensive 3' exonuclease activity that remove nucleotides from the 3' terminus and subsequently fills in with the radioactive nucleotides provided. This results in both ends becoming equally labeled. To achieve unique end-labeling it was necessary to select an enzyme lacking exonuclease activity. For this reason, end-labeling reactions were performed with AMV reverse transcriptase.

Linear bacteriophage λ DNA (7.5 μ g) was incubated in 100 mM tris-HCl, pH 8.3, 140 mM KCl, 10 mM MgCl₂, 28 mM β -mercapthoethanol, and 5 units of AMV reverse transcriptase. For unique labeling of the right end, 0.6 mM dCTP and 2.5 μ l of [α - 32 P]dGTP were added to the 50 μ l final reaction volume (Fig. 2.10). For left end labeling, 2.5 μ l of [α - 32 P]dATP was added. The unincorporated mononucleotides were removed by G-50 Sephadex spin column chromatography preequilibrated in H₂O. Radioactive incorporation

was quantitated by scintillation counting and the DNA diluted to 5000 cpm per microliter. The quality of the DNA was improved and the shelf-life extended if the DNA was stored in a buffered solution (250 mM NaCl, 62 mM tris-acetate, pH 6.6-7.4).

Cleavage Reactions. Affinity cleaving reaction conditions are described in the figure legends. Cleavage efficiencies were determined either by laser densitometry of the autoradiograms or radioactive scintillation counting of the excised bands. Efficiencies were expressed as a ratio of radioactive emissions in the product band to the sum of emissions in the product and uncut substrate bands.

References

1. P. B. Dervan, Design of Sequence-Specific DNA-Binding Molecules *Science* **232**, 464-468 (1986).
2. P. B. Dervan, Sequence Specific Recognition of Double Helical DNA. A Synthetic Approach in *Nucleic Acids and Molecular Biology* F. Eckstein, D. M. J. Lilley, Eds. (Springer-Verlag, Berlin Heidelberg, 1988), pp. 49-64.
3. H. E. Moser, P. B. Dervan, Sequence-Specific Cleavage of Double Helical DNA by Triple Helix Formation *Science* **238**, 645-650 (1987).
4. S. A. Strobel, H. E. Moser, P. B. Dervan, Double-Strand Cleavage of Genomic DNA at a Single Site by Triple Helix Formation *J. Amer. Chem. Soc.* **110**, 7927-7929 (1988).
5. J. Dreyer, P. B. Dervan, Sequence-specific Cleavage of Single-stranded DNA: Oligodeoxynucleotide-EDTA•Fe(II) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 968-972 (1985).
6. F. Sanger, A. R. Coulson, G. F. Hong, D. F. Hill, G. B. Petersen, Nucleotide Sequence of Bacteriophage λ DNA *J. Mol. Biol.* **162**, 729-773 (1982).

7. S. A. Strobel, P. B. Dervan, Cooperative Site Specific Binding of Oligonucleotides to Duplex DNA *J. Amer. Chem. Soc.* **111**, 7286-7287 (1989).
8. K. Dennison-Thompson, D. D. Moore, K. E. Kruger, M. E. Furth, F. R. Blattner, Physical Structure of the Replication Origin of Bacteriophage Lambda *Science* **198**, 1051-1056 (1977).
9. M. E. Furth, F. R. Blattner, C. McLeester, W. F. Dove, Genetic Structure of the Replication Origin of Bacteriophage Lambda *Science* **198**, 1046-1051 (1977).
10. M. E. Furth, S. H. Wickner, Lambda DNA Replication in *Lambda II* R. W. Hendrix, J. W. Roberts, F. W. Stahl, R. A. Weisberg, Eds. (Cold Spring Harbor Press, New York, 1983), pp. 145-173.
11. S. Wickner, K. McKenney, Deletion Analysis of the DNA Sequence Required for the *in Vitro* Initiation of Replication of Bacteriophage λ *J. Biol. Chem.* **262**, 13163-13167 (1987).
12. K. Zahn, F. R. Blattner, Sequence-Induced DNA Curvature at the Bacteriophage λ Origin of Replication *Nature* **317**, 451-453 (1985).
13. M. Dodson, H. Echols, S. Wickner, C. Alfano, K. Mensa-Wilmot, B. Gomes, J. LeBowitz, J. D. Roberts, R. McMacken, Specialized Nucleoprotein Structures at the Origin of Replication of Bacteriophage λ : Localized Unwinding of Duplex DNA by a Six-Protein Reaction *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7638-7642 (1986).
14. M. Schnos, K. Zahn, R. B. Inman, F. R. Blattner, Initiation Protein Induced Helix Destabilization at the λ Origin: A Prepriming Step in DNA Replication *Cell* **52**, 385-395 (1988).
15. R. McMacken, R. Jordon, S. J. Um, L. Huang, B. Learn, K. Carroll, D. S. Sampath, Molecular Mechanisms in the Initiation of Bacteriophage λ DNA Replication A1064 (1991).
16. K. Zahn, F. R. Blattner, Binding and Bending of the λ Replication Origin by the Phage O Protein *EMBO J.* **4**, 3605-3616 (1985).
17. K. Zahn, F. R. Blattner, Direct Evidence for DNA Bending at the Lambda Replication Origin *Science* **236**, 416-422 (1987).

18. V. I. Lyamichev, S. M. Mirkin, M. D. Frank-Kamenetskii, Structures of Homopurine-Homopyrimidine Tract in Superhelical DNA *J. Biomole. Struct. Dyn.* **3**, 667-669 (1986).
19. H. Htun, J. E. Dahlberg, Single Strands, Triple Strands, and Kinks in H-DNA *Science* **241**, 1791-1796 (1988).
20. B. H. Johnston, The S1-Sensitive Form of $d(C-T)_n \cdot d(A-G)_n$: Chemical Evidence for a Three-Stranded Structure in Plasmids *Science* **241**, 1800-1804 (1988).
21. J. C. Hanvey, M. Shimizu, R. D. Wells, Intramolecular DNA Triplexes in Supercoiled Plasmids *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6292-6296 (1988).
22. S. M. Mirkin, V. I. Lyamichev, K. N. Drushlyak, V. N. Dobrynin, S. A. Filippov, M. D. Frank-Kamenetskii, DNA H form Requires a Homopurine-Homopyrimidine Mirror Repeat *Nature* **330**, 495-497 (1987).
23. O. N. Voloshin, S. M. Mirkin, V. I. Lyamichev, B. P. Belotserkovskii, M. D. Frank-Kamenetskii, Chemical Probing of Homopurine-Homopyrimidine Mirror Repeats in Supercoiled DNA *Nature* **333**, 475-476 (1988).
24. T. J. Povsic, P. B. Dervan, Triple Helix Formation by Oligonucleotides on DNA Extended to the Physiological pH Range *J. Am. Chem. Soc.* **111**, 3059-3061 (1989).
25. S. A. Strobel, P. B. Dervan, Single-site Enzymatic Cleavage of Yeast Genomic DNA Mediated by Triple Helix Formation *Nature* **350**, 172-174 (1991).
26. F. Hutchinson, The Lesions Produced by Ultraviolet Light in DNA Containing 5-Bromouracil *Quart. Rev. Biophys.* **6**, 201-246 (1973).
27. K. M. Swierkowska, J. K. Jasinska, J. A. Steffen, 5-Ethyl-2'-Deoxyuridine: Evidence for Incorporation into DNA and Evaluation of Biological Properties in Lymphocyte Cultures Grown Under Conditions of Amethopterin-Imposed Thymidine Deficiency *Biochem. Pharmacol.* **22**, 85-93 (1973).
28. E. De Clercq, D. Shugar, Antiviral Activity of 5-Ethyl Pyrimidine Deoxynucleosides *Biochem. Pharm.* **24**, 1073-1078 (1975).
29. P. D. Lawley, P. Brooks, Ionization of DNA Bases or Base Analogues as a Possible Explanation of Mutagenesis with Special Reference to 5-Bromodeoxyuridine *J. Mol. Biol.* **4**, 216-219 (1962).

30. P. H. Driggers, K. L. Beattie, Effect of pH on the Base-Mispairing Properties of 5-Bromouracil during DNA Synthesis *Biochem.* **27**, 1729-1735 (1988).
31. E. Freese, The Specific Mutagenic Effect of Base Analogues on Phage T4 *J. Mol. Biol.* **1**, 87-105 (1959).
32. A. R. Katritsky, A. J. Waring, Tautomeric Azines. Part I. The Tautomerism of 1-Methyluracil and 5-Bromo-1-Methyluracil *J. Chem. Soc.* **1962**, 1540-1544 (1962).
33. T. J. Povsic, Oligonucleotide Directed Sequence Specific Recognition and Alkylation of Double Helical DNA by Triple Helix Formation (California Institute of Technology, 1992).
34. L. J. Maher, B. Wold, P. B. Dervan, Inhibition of DNA Binding Proteins by Oligonucleotide-Directed Triple Helix Formation *Science* **245**, 725-730 (1989).
35. L. J. Maher, P. B. Dervan, B. Wold, Kinetic Analysis of Oligodeoxyribonucleotide-Directed Triple-Helix Formation on DNA *Biochemistry* **29**, 8820-8826 (1990).
36. T. D. Tullius, B. A. Dombroski, Iron(II) EDTA Used to Measure the Helical Twist Along Any DNA Molecule *Science* **230**, 679-681 (1985).
37. T. D. Tullius, B. A. Dombroski, Hydroxyl Radical "Footprinting": High-resolution Information About DNA-protein Contacts and Application to λ Repressor and Cro Protein *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5469-5473 (1986).
38. L. C. Griffin, Oligonucleotide-Directed Cleavage of Single- and Double-Stranded DNA by Double and Triple Helix Formation (California Institute of Technology, 1990).
39. S. B. Lin, K. R. Blake, P. S. Miller, P. O. P. Ts'o, Use of EDTA Derivatization to Characterize Interactions between Oligodeoxyribonucleoside Methylphosphonates and Nucleic Acids *Biochemistry* **28**, 1054-1061 (1989).
40. M. D. Matteucci, M. H. Caruthers, Synthesis of Deoxyoligonucleoties on a Polymer Support *J. Amer. Chem. Soc.* **103**, 3185-3191 (1981).

41. T. Atkinson, M. Smith, Solid Phase Synthesis of Oligodeoxyribonucleotides by the Phosphite-Triester Method in *Oligonucleotide Synthesis-A Practical Approach* M. G. Gait, Eds. (IRL Press, Oxford, 1985), pp. 35-81.
42. P. A. Beal, Thesis (California Institute of Technology, 1994).
43. P. A. Beal, P. B. Dervan, Second Structural Motif for Recognition of DNA by Oligonucleotide-Directed Triple Helix Formation *Science* **251**, 1360-1363 (1991).
44. D. Thiele, W. Guschlbauer, Protonated Polynucleotide Structures. IX. Disproportionation of Poly (G)•Poly (C) in Acid Medium *Biopolymers* **10**, 143-157 (1971).
45. J. S. Lee, D. A. Johnson, A. R. Morgan, Complexes Formed by (Pyrimidine)_n•(Purine)_n DNAs on Lowering the pH are Three-Stranded *Nuc. Acids Res.* **6**, 3073-3091 (1979).
46. J. S. Lee, M. L. Woodsworth, L. J. P. Latimer, A. R. Morgan, Poly(pyrimidine)•poly(purine) Synthetic DNAs Containing 5-methylcytosine Form Stable Triplexes at Neutral pH *Nuc. Acids Res.* **12**, 6603-6614 (1984).
47. V. I. Lyamichev, S. M. Mirkin, M. D. Frank-Kamenetskii, C. R. Cantor, A Stable Complex Between Homopyrimidine Oligomers and the Homologous Regions of Duplex DNAs *Nuc. Acids Res.* **16**, 2165-2178 (1988).
48. S. A. Strobel, P. B. Dervan, Site-Specific Cleavage of a Yeast Chromosome by Oligonucleotide-Directed Triple-Helix Formation *Science* **249**, 73-75 (1990).
49. W. D. Wilson, Cooperative Effects in Drug-DNA Interactions *Prog. Drug Res.* **31**, 193-221 (1987).
50. M. Ptashne, How Eukaryotic Transcriptional Activators Work *Nature* **335**, 683- (1988).
51. S. Adhya, Multipartite Genetic Control Elements: Communication by DNA Loop *Ann. Rev. Genet.* **1989**, 227-250 (1989).
52. E. Giniger, M. Ptashne, Cooperative DNA Binding of the Yeast Transcriptional Activator GAL4 *Proc. Natl. Acad. Sci. U.S.A.* **85**, 382-386 (1988).

53. A. Hirano, T. Wong, Functional Interaction between Transcriptional Elements in the Long Terminal Repeat of Reticuloendotheliosis Virus: Cooperative DNA Binding of Promoter- and Enhancer-Specific Factors *Mole. Cell. Biol.* **8**, 5232-5244 (1988).
54. A. Hochschild, M. Ptashne, Cooperative Binding of λ Repressors to Sites Separated by Integral Turns of the DNA Helix *Cell* **44**, 681-687 (1986).
55. A. D. Johnson, B. J. Meyer, M. Ptashne, Interactions Between DNA-Bound Repressors Govern Regulation by the λ Phage Repressor *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5061-5065 (1979).
56. J. D. McGhee, P. H. von Hippel, Theoretical Aspects of DNA-Protein Interactions: Co-operative and Non-co-operative Binding of Large Ligands to a One-dimensional Homogeneous Lattice *J. Mol. Biol.* **86**, 469-489 (1974).
57. L. Poellinger, B. K. Yoza, R. G. Roeder, Functional Cooperativity between Protein Molecules Bound at Two Distinct Sequence Elements of the Immunoglobulin Heavy-Chain Promoter *Nature* **337**, 573-576 (1989).
58. Y. L. Ren, S. Garges, S. Adhya, J. S. Krakow, Cooperative DNA Binding of Heterologous Proteins: Evidence for Contact Between the Cyclic AMP Receptor Protein and RNA Polymerase *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4138-4142 (1988).
59. R. Schule, M. Muller, H. Otsuka-Murakami, R. Renkawitz, Cooperativity of the Glucocorticoid Receptor and the CACCC-Box Binding Factor *Nature* **332**, 87-90 (1988).
60. D. F. Senear, M. Brenowitz, M. A. Shea, G. K. Ackers, Energetics of Cooperative Protein-DNA Interactions: Comparison between Quantitative Deoxyribonuclease Footprint Titration and Filter Binding *Biochemistry* **25**, 7344-7354 (1986).
61. W. S. Wade, Sequence Specific Complexation of B DNA at Sites Containing G,C Base Pairs (California Institute of Technology, 1989).
62. M. Hogan, N. Dattagupta, D. M. Crothers, Transmission of Allosteric Effects in DNA *Nature* **278**, 521-524 (1979).
63. G. T. Walker, M. P. Stone, T. R. Krugh, Ethidium Binding to Left-Handed (Z) DNAs Results in Regions of Right-Handed DNA at the Intercalation Site *Biochem.* **24**, 7462-7471 (1985).

64. J. B. Chaires, Long-Range Allosteric Effects on the B to Z Equilibrium by Daunomycin *Biochemistry* **24**, 7479-7486 (1985).
65. G. T. Walker, M. P. Stone, T. R. Krugh, Interaction of Drugs with Z-DNA: Cooperative Binding of Actinomycin D or Actinomine to the Left-Handed Forms of Poly(dG-dC)•Poly(dG-dC) and Poly(dG-m⁵dC)•Poly(dG-m⁵dC) Reverses the Conformation of the Helix *Biochem.* **24**, 7471-7479 (1985).
66. M. W. Springgate, D. Poland, Cooperative and Thermodynamic Parameters for Oligoinosinate-Polycytidylate Complexes *Biopolymers* **12**, 2241-2260 (1973).
67. U. Asseline, M. Delarue, G. Lancelot, F. Toulme, N. T. Thuong, T. Montenay-Garestier, C. Helene, Nucleic Acid-Binding Molecules with High Affinity and Base Sequence Specificity: Intercalating Agents Covalently Linked to Oligodeoxynucleotides *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3297-3301 (1984).
68. L. J. Maher, B. J. Dolnick, Specific Hybridization Arrest of Dihydrofolate Reductase mRNA *in vitro* Using Anti-sense RNA or Anti-sense Oligonucleotides *Arch. Biochem. Biophys.* **253**, 214-220 (1987).
69. D. Horne, P. B. Dervan, Effects of an Abasic Site on Triple Helix Formation Characterized by Affinity Cleaving *Nuc. Acids Res.* (In Press).
70. M. Distefano, P. B. Dervan, *Unpublished Results* (1991).
71. J. M. L. Pieters, R. M. W. Mans, H. van den Elst, G. A. van der Marel, J. H. van Boom, C. Altona, Conformational and Thermodynamic Consequences of the Introduction of a Nick in Duplexed DNA Fragments: an NMR Study Augmented by Biochemical Experiments *Nuc. Acids Res.* **17**, 4551-4565 (1989).
72. J. Aymani, M. Coll, G. A. van der Marel, J. H. van Boom, A. H. J. Wang, A. Rich, Molecular Structure of Nicked DNA: A Substrate for DNA Repair Enzymes *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2526-2530 (1990).
73. E. Wunderlich, M. Distefano, P. B. Dervan, Quantitation of End-End and Minihelix Cooperativity in the Triple Helix Motif by Restriction Enzyme Inhibition *Unpublished Results* (1991).
74. M. D. Distefano, J. A. Shin, P. B. Dervan, Cooperative Binding of Oligonucleotides to DNA by Triple Helix Formation: Dimerization via Watson-Crick Hydrogen Bonds *J. Amer. Chem. Soc.* (1991).

75. K. S. Gates, P. B. Dervan, Metalloregulated Binding of an Oligonucleotide to Double Helical DNA via Triple-Helix Formation *J. Amer. Chem. Soc.* (In Preparation).

Chapter III

Site Specific Cleavage of Yeast Genomic DNA Mediated by Triple Helix Formation

Site specific cleavage of bacteriophage λ DNA (10^4 bp) demonstrated the utility of using oligonucleotide directed triple helix formation for genomic cleavage (1). The next incremental step toward site specific cleavage of a human chromosome (10^9 bp) was to demonstrate that an oligonucleotide was capable of mediating target site cleavage within yeast genomic DNA (10^7 bp). The *Saccharomyces cerevisiae* genome contains more than 14 megabase pairs (Mb) of DNA divided among 16 autonomously segregating chromosomes (2). The chromosomes range in size from 220 kb to 2.5 Mb and are readily resolvable by pulsed-field gel electrophoresis (3-8). Chromosome III, one of the smallest yeast chromosomes (340 kb), was chosen as a target because of its ease of manipulation and resolution. Yeast is an ideal organism for genetic manipulations, and many genes and sequences have been mapped to specific chromosomal locations (2, 9). One such gene is *LEU2* which maps to the short arm of chromosome III (3p) near the centromere (9), and encodes for β -isopropylmalate dehydrogenase, an enzyme that catalyzes the third step in leucine biosynthesis (10-12). *LEU2* can be used to specifically insert a triple helix target sequence by homologous recombination at a defined position in a yeast chromosome and select for the insertion by screening cells for the ability to synthesize leucine .

This chapter describes the site specific cleavage of yeast chromosome III by oligonucleotide directed triple helix formation. It is divided into two sections based on the method of cleavage. The first describes the site specific

cleavage of yeast chromosome III by oligonucleotides containing the affinity cleaving agent EDTA•Fe(II) (T*) (13). It includes the use of specific, base modified, and nonspecific oligonucleotides to cut yeast chromosome III by affinity cleaving, and an analysis of the specificity of triple helix formation as a function of pH and oligonucleotide composition. The major findings of this work were published in *Science* 249, 73-75 (1990) (14). The second section describes single site cleavage of the yeast genome by a modification of the Achilles heel technique for enzymatic cleavage, and analyzes the effects of oligonucleotide composition and pH on cleavage specificity by this technique. The results were reported in two publications. The report in *Nature* 350, 172-174 (1991) emphasized the experimental result (15), while the paper to appear in *Methods in Enzymology* (in press) provides a detailed description of the protocol (16). This chapter also contains a methods section that provides a complete description of plasmid construction, chromosomal recombination, and polymerase chain reaction (PCR) protocols used to generate and verify the yeast strains used as genomic substrates in these reactions.

Reprint Series
6 July 1990, Volume 249, pp. 73-75

Part I

**Site-Specific Cleavage of a Yeast Chromosome by
Oligonucleotide-Directed Triple-Helix Formation**

SCOTT A. STROBEL AND PETER B. DERVAN*

Site-Specific Cleavage of a Yeast Chromosome by Oligonucleotide-Directed Triple-Helix Formation

SCOTT A. STROBEL AND PETER B. DERVAN*

Oligonucleotides equipped with EDTA-Fe can bind specifically to duplex DNA by triple-helix formation and produce double-strand cleavage at binding sites greater than 12 base pairs in size. To demonstrate that oligonucleotide-directed triple-helix formation is a viable chemical approach for the site-specific cleavage of large genomic DNA, an oligonucleotide with EDTA-Fe at the 5' and 3' ends was targeted to a 20-base pair sequence in the 340-kilobase pair chromosome III of *Saccharomyces cerevisiae*. Double-strand cleavage products of the correct size and location were observed, indicating that the oligonucleotide bound and cleaved the target site among almost 14 megabase pairs of DNA. Because oligonucleotide-directed triple-helix formation has the potential to be a general solution for DNA recognition, this result has implications for physical mapping of chromosomes.

TECHNIQUES FOR THE SITE-SPECIFIC cleavage of double-stranded DNA are vital to chromosomal mapping, gene isolation, and DNA sequencing (1, 2). Restriction endonucleases with 4- to 6-base pair (bp) binding sites cleave too frequently for many chromosomal DNA manipulations (3). Rare-cutting restriction enzymes with 8-bp specificities have found widespread use in genetic mapping; however, these enzymes are few in number, are limited to the recognition of CpG-rich sequences, and cleave at sites that tend to be highly clustered (4). Combinations of methylases and restriction enzymes that require methylated sequences can produce cleavage specificities of 8 to 12 bp (5). Transient methylase protection can be induced by DNA binding proteins that recognize sequences with overlapping restriction-methylation sites; restriction enzyme digestion then produces specific cleavage at the protein binding site (6). Recently, endonucleases encoded by group I introns have been discovered that might have greater than 12 bp specificity (7). Unfortunately, none of these strategies can be generalized to recognize and cleave at any of the large number of unique sequences contained in human DNA.

Pyrimidine oligonucleotides bind specifically to purine sequences in duplex DNA to form a local triple-helix structure (8-12). The oligonucleotide binds in the major groove parallel to the Watson-Crick purine strand by Hoogsteen hydrogen bonding (8-12). Triple-helix specificity is derived from thymine (T) binding to adenine-thymine base pairs (T·A·T base triplet) and protonated cytosine (C+) binding to guanine-cyto-

sine base pairs (C + GC base triplet) (8-15). Guanine recognition of thymine-adenine base pairs (G·T·A base triplet) within the pyrimidine triple-helix motif (9) and recognition of (purine)_n(pyrimidine)_m type sequences by alternate strand triple-helix formation (10) have extended recognition of duplex DNA to a wide class of mixed purine-pyrimidine sequences (16). Oligonucleotides 15 to 20 bases in length equipped with an EDTA-Fe moiety produce sequence-specific double-strand breaks with efficiencies ranging from 5 to 25% at their target sites within genomes as large as that of bacteriophage λ (48.5 kbp) (8-10). In order to determine if this specificity can be achieved in chromosomal DNA, a triple-helix target site, 5'-A₂GA₄GA₂GA₃GA₅-3', was inserted proximal to the *LEU2* gene on the short arm of the 340-kb chromosome III of *Saccharomyces cerevisiae* (17-25) by homologous recombination (Fig. 1). The ge-

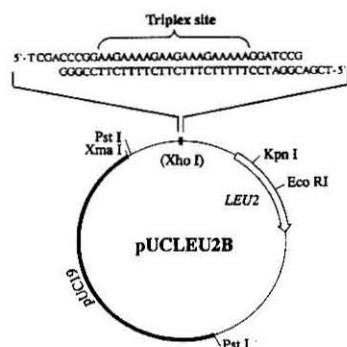


Fig. 1. Schematic diagram of pUCLEU2B constructed by insertion of the Pst I-Xma I 4.0-kb *LEU2* fragment from YEp13 into pUC19 by standard procedures (19). Complementary oligonucleotides containing a homopurine sequence were ligated into the unique Xho I site, upstream of the *LEU2* gene.

netic map location of the *LEU2* locus indicates that double-strand cleavage at the 20-bp target site should produce two fragments, approximately 110 \pm 10 and 230 \pm 10 kb in size (26) (Fig. 2). We report the site-specific cleavage at this genetically engineered sequence on chromosome III by an oligonucleotide-(EDTA-Fe)₂.

A 20-base pyrimidine oligonucleotide, 5'-T·TCT₄CT₂CT₃CT₄T*-3', with thymidine EDTA (T*) (27) at the 5' and 3' termini, was synthesized by automated methods beginning with 5'-O-DMT-thymidine-EDTA-triethylster 3'-succinyl control pore glass as the solid support (DMT, 4,4'-dimethoxytrityl). Cleavage reactions were performed on yeast transformants SEY6210 (- target site) and SEY6210B (+ target site) (28). Chromosomal DNA embedded in an agarose plug was equilibrated with oligonucleotide-(EDTA-Fe)₂ to facilitate diffusion into the agarose and triple-helix formation (pH 7.2, 22°C). The cleavage reaction was initiated by addition of dithiothreitol (DTT). To improve the cleavage yield, a second cleavage cycle was performed by disrupting the triplex at conditions of high pH (8.5), reequilibrating the plug in a triplex-compat-

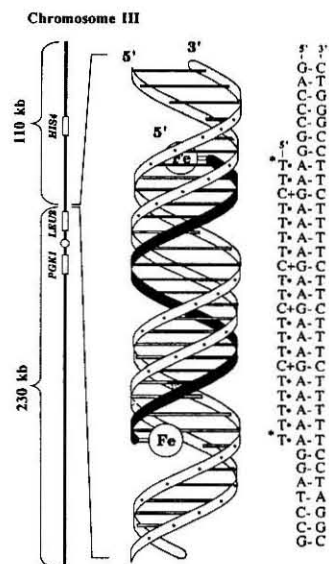


Fig. 2. (Left) Genetic map of *S. cerevisiae* strain SEY6210B (+ target site) chromosome III. The locations of *HIS4*, *PGK1*, and *LEU2* loci (boxes), the centromere (circle), and the triple-helix target site are indicated. The sizes of the cleavage products (based upon genetic map distances and experimental results) are shown. (Right) Schematic diagram of the triple-helix complex. The pyrimidine oligonucleotide with EDTA-Fe at the 5' and 3' termini is bound in the major groove parallel to the purine strand.

Arnold and Mabel Beckman Laboratories of Chemical Synthesis, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125.

*To whom correspondence should be addressed.

ible buffer, and repeating the reaction with fresh reagents. The chromosomes were separated by pulsed-field gel electrophoresis and detected by ethidium bromide staining (Fig. 3, A and B). Cleavage products were detected by DNA blotting with chromosome III-specific probes (Fig. 3, C and D).

The *HIS4* (29) and *PGK1* (30) genes are located on the short and long arms of chromosome III, respectively (Fig. 2). DNA hybridization of the resolved cleavage products (Fig. 3A) with a radiolabeled *HIS4* probe revealed a 110 ± 10 kb fragment present only in the yeast strain containing the engineered target site (SEY6210B) (Fig. 3C, lanes 3 and 4). Hybridization with a radiolabeled *PGK1* probe revealed a second unique fragment 230 ± 10 kb in size (Fig. 3D, lanes 3 and 4). The extent of double-strand cleavage at the target site was estimated at 6% by densitometry. The observed fragment sizes are consistent with those estimated from the genetic map (26). Thus, after searching through almost 14 megabase pairs of yeast DNA, the oligonucleotide

bound and cleaved specifically at the 20-bp target site while leaving the other chromosomes largely intact (Fig. 3B).

The sequence specificity of pyrimidine oligonucleotides for local triple-helix formation on duplex DNA is dependent upon pH, temperature, and organic cosolvents (8). Under conditions of lower pH, lower temperature, or added ethanol, oligonucleotides have been observed to bind to sites that are in significant but not perfect match with the target-site sequence (8). Because the complete sequence of the yeast genome is not yet available, the location and number of secondary binding sites on chromosome III could not be predicted a priori. Interestingly, one major (300 ± 10 kb) and three minor (190, 210, and 240 ± 10 kb) secondary cleavage fragments were detected on chromosome III at pH 7.2 (Fig. 3, C and D, lanes 3 and 4) (31). The appearance of the three minor fragments (190, 210, and 240 kb) upon hybridization with the flanking markers *HIS4* and *PGK1* indicates that the minor secondary cleavage sites are found on the long arm of chromosome III, distal to the engineered target site. The major secondary cleavage site (300 kb) was not flanked by the markers, but must map to within 40 kb of a chromosome III telomere.

The extent of sequence similarity of the secondary sites to the target site can be estimated by examining the cleavage pattern as a function of increasing pH. The cleavage products were examined over the pH range 7.2 to 7.8 (Fig. 3E). The 190- and 210-kb bands were not observed above pH 7.4

(lanes 7 to 10), whereas raising the pH above 7.6 eliminated the 240-kb fragment (lanes 9 and 10). The 300-kb band and the fragment corresponding to the designed target site were still observed at pH 7.8 (lanes 9 and 10) though at lower cleavage efficiencies. This suggests that the order of sequence similarity of the different sites with the target site are $300 > 240 > 210, 190$ kb.

A chemical approach for the site-specific cleavage of intact chromosomes at 12- to 20-bp sequences might assist the large effort being directed toward mapping genomic DNA. For an unambiguous test of site-specific cleavage on chromosomal DNA by oligonucleotide-directed triple-helix formation, a target site of known sequence and approximate physical location was chosen for this experiment. However, the ability of oligonucleotide-directed triple-helix formation to recognize a wide variety of purine and mixed purine-pyrimidine sequences (16) could allow the orchestrated cleavage of large genomic DNA at any genetic marker for which some sequence information is known.

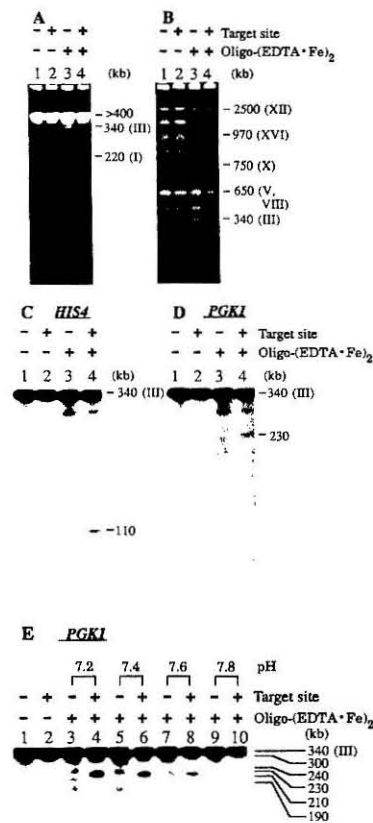


Fig. 3. Site-specific cleavage of yeast chromosomes. Lanes 1 and 2 (all gels): SEY6210 (- target site) and SEY6210B (+ target site) chromosomal DNA unreacted controls, respectively. Lanes 3 and 4 (all gels): SEY6210 and SEY6210B chromosomal DNA, respectively, after reaction with oligonucleotide-(EDTA-Fe)₂. (A) Separation of yeast chromosomes less than 400 kb in size by pulsed-field gel electrophoresis on a Bio-Rad CHEF system. Pulse times were ramped from 10 to 20 s during a 24-hour period (14°C and 200 V). Chromosomal DNA was detected by ethidium bromide staining. Fragment sizes were estimated by comparison to bacteriophage λ concatemers. (B) Separation of all yeast chromosomes with tentative assignments. A 60-s pulse time for 16 hours was followed by 90-s pulses for 8 hours (14°C, 200 V). Sizes were estimated by comparison to YNN295 chromosomal DNA (24). (C) DNA blot hybridization of reactions shown in (A) with a 250-bp *HIS4* fragment labeled with ³²P by random priming (19). The DNA blot transfer and hybridization were performed by standard procedures (19). The cleavage products were visualized by autoradiography and quantitated by laser densitometry. (D) DNA blot hybridization as in (C) except a 1.3-kb marker from the promoter region of *PGK1* was used for hybridization (E). pH profile of cleavage products hybridized with *PGK1* marker.

REFERENCES AND NOTES

- H. O. Smith, *Science* **205**, 455 (1979).
- P. Modrich, *Crit. Rev. Biochem.* **13**, 287 (1982).
- P. B. Dervan, *Science* **232**, 464 (1986); in *Nucleic Acids and Molecular Biology*, F. Eckstein and D. M. J. Lilley, Eds. (Springer-Verlag, Heidelberg, 1988), vol. 2, pp. 49-64.
- D. P. Barlow and H. Lehrach, *Trends Genet.* **3**, 167 (1987).
- M. McClelland, L. G. Kessler, M. Bitner, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 983 (1984); M. McClelland, M. Nelson, C. R. Cantor, *Nucleic Acids Res.* **13**, 7171 (1985); M. D. Weil and M. McClelland, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 51 (1989); Y. Patel, E. Van Cott, G. G. Wilson, M. McClelland, *Nucleic Acids Res.*, in press.
- M. Koob, E. Grimes, W. Szybalski, *Science* **241**, 1084 (1988).
- P. S. Perlman and R. A. Butow, *ibid.* **246**, 1106 (1989).
- H. E. Moser and P. B. Dervan, *ibid.* **238**, 645 (1987); S. A. Strobel, H. E. Moser, P. B. Dervan, *J. Am. Chem. Soc.* **110**, 7927 (1988); T. J. Povsic and P. B. Dervan, *ibid.* **111**, 3059 (1989); L. J. Maher III, B. Wold, P. B. Dervan, *Science* **245**, 725 (1989); S. A. Strobel and P. B. Dervan, *J. Am. Chem. Soc.* **111**, 7286 (1989).
- L. C. Griffin and P. B. Dervan, *Science* **245**, 967 (1989).
- D. A. Horne and P. B. Dervan, *J. Am. Chem. Soc.* **112**, 2435 (1990).
- T. Le Doan et al., *Nucleic Acids Res.* **15**, 7749 (1987); V. I. Lyamichev, S. M. Mirkin, M. D. Frank-Kamenetskii, C. R. Cantor, *ibid.* **16**, 2165 (1988); D. Praseuth et al., *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1349 (1988); J. C. Francois, T. Saison-Behmoaras, M. Chassinol, N. T. Thuong, C. Héline, *J. Biol. Chem.* **264**, 5891 (1989); J. C. Francois, T. Saison-Behmoaras, N. T. Thuong, C. Héline, *Biochemistry* **28**, 9617 (1989); J. S. Sun et al., *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9198 (1989).
- P. Rajagopal and J. Feigon, *Nature* **239**, 637 (1989); C. de los Santos, M. Rosen, D. Patel, *Biochemistry* **28**, 7282 (1989).
- For triple-stranded polynucleotides, see: G. Felsenfeld, D. R. Davies, A. Rich, *J. Am. Chem. Soc.* **79**,

- 2023 (1957); K. Hoogsteen, *Acta Crystallogr.* **12**, 822 (1959); M. N. Lipsett, *Biochem. Biophys. Res. Commun.* **11**, 224 (1963); *J. Biol. Chem.* **239**, 1256 (1964); A. M. Michelson, J. Massoulié, W. Guschlbauer, *Prog. Nucleic Acids Res. Mol. Biol.* **6**, 83 (1967); G. Felsenfeld and H. T. Miles, *Annu. Rev. Biochem.* **36**, 407 (1967); F. B. Howard, J. Frazier, M. N. Lipsett, H. T. Miles, *Biochem. Biophys. Res. Commun.* **17**, 93 (1964); J. H. Miller and J. M. Sobell, *Proc. Natl. Acad. Sci. U.S.A.* **55**, 1201 (1966); A. R. Morgan and R. D. Wells, *J. Mol. Biol.* **37**, 63 (1968); J. S. Lee, D. A. Johnson, A. R. Morgan, *Nucleic Acids Res.* **6**, 3073 (1979).
14. For intramolecular triplexes (H-form DNA), see: S. M. Mirkin *et al.*, *Nature* **330**, 495 (1987); O. N. Voloshin, S. M. Mirkin, V. I. Lyamichev, B. P. Belotserkovskii, M. D. Frank-Kamenetskii, *ibid.* **333**, 475 (1988); H. Htin and J. E. Dahlberg, *Science* **241**, 1791 (1988); B. H. Johnston, *ibid.*, p. 1800; Y. Kohwi and T. Kohwi-Shigematsu, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3781 (1988); R. D. Wells, D. A. Collier, J. C. Hanvey, M. Shimizu, F. Wohlrab, *FASEB J.* **2**, 2939 (1988).
 15. For oligonucleotide recognition extended to purine motifs (G-GC base triplets), see: M. Cooney, G. Czernuszewicz, E. H. Postel, S. J. Flint, M. E. Hogan, *Science* **241**, 456 (1988).
 16. Recognition sites 16 bp in size have 2,147,516,416 unique sequences which is the size range of the human genome (3×10^9 bp) (3). In a formal sense, pyrimidine oligonucleotide-directed triple-helix formation utilizing T:AT and C + GC base triplets could recognize 65,536 homopurine sites. The ability to recognize one T:A base pair by use of a single G:TA base triplet in a 16-base sequence (9) yields an additional 524,288 binding sites. Alternate strand triple-helix formation (10) of the type (purine)_n-NN(pyrimidine)_m ($n = 1$ to 7 and $n + m = 14$) affords 967,044 additional sequences. Thus the total number of 16-bp sites potentially recognized by current pyrimidine triple-helix motifs is 1,556,868. This is approximately one site per 2000 bp in the human genome. Sequence composition effect should be heeded when considering these values.
 17. An oligonucleotide duplex containing the 20-bp target site was ligated 650-bp upstream of the *LEU2* gene at the unique Xho I site of the yeast shuttle vector YEp13 (18). The orientation of the insert was determined by sequencing from a *LEU2*-specific primer. The 4.0-kb Pst I-Xma I fragment containing *LEU2* and flanking sequences was subcloned (19) into pUC19 by a three-piece ligation to generate pUCLEU2B (Fig. 1). Competent haploid *S. cerevisiae* cells (SEY6210 *leu2*) (20) were transformed (21) with Pst I-linearized pUCLEU2B DNA and recombinants were selected on leucine-deficient minimal media. Chromosomal DNA from recombinant colonies was prepared from a log phase yeast culture by spheroplasting with Zymolyase in 0.9 M sorbitol at 37°C (pH 5.6) followed by sarcosyl and Proteinase K treatment in 0.5% low melting point agarose and 0.5 M EDTA at 50°C (22). Insertion of the Pst I *LEU2* fragment into chromosome III was confirmed by pulsed-field gel electrophoretic separation of the yeast chromosomes (23) followed by DNA blotting (19, 24) with the random primer ³²P-labeled Kpn I-EcoR I fragment from the *LEU2* gene (Fig. 1). The presence of the triple-helix target site in the yeast construct was verified by the polymerase chain reaction (25) with a *LEU2*-specific oligonucleotide and copies of the inserted oligonucleotides as primers for amplification.
 18. J. R. Broach *et al.*, *Gene* **8**, 121 (1979); A. Andreadis, Y. P. Hsu, G. B. Kohlhaw, P. Schimmel, *Cell* **31**, 319 (1982); D. A. Fischhoff, R. H. Waterston, M. V. Olson, *Gene* **27**, 239 (1984).
 19. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).
 20. For complete genotype, see J. S. Robinson *et al.*, *Mol. Cell. Biol.* **8**, 4936 (1988).
 21. A. Hinnen, J. B. Hicks, G. R. Fink, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1929 (1978); H. Ito, Y. Fukuda, K. Murata, A. Kimura, *J. Bacteriol.* **153**, 163 (1983).
 22. C. L. Smith, S. R. Kleo, C. R. Cantor, in *Genome Analysis: A Practical Approach*, K. Davies, Ed. (IRL Press, Oxford, 1988), pp. 41-72.
 23. D. C. Schwartz and C. R. Cantor, *Cell* **37**, 67 (1984); G. F. Carle and M. V. Olson, *Nucleic Acids Res.* **12**, 5647 (1984); G. F. Carle, M. Frank, M. V. Olson, *Science* **232**, 65 (1986); S. M. Clark, E. Lai, B. W. Birren, L. Hood, *ibid.* **241**, 1203 (1988); B. W. Birren, E. Lai, S. M. Clark, L. Hood, M. I. Simon, *Nucleic Acids Res.* **16**, 7563 (1988).
 24. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
 25. R. K. Saiki *et al.*, *Science* **230**, 1350 (1985); R. K. Saiki *et al.*, *ibid.* **239**, 487 (1988).
 26. Chromosome III is approximately 140 centimorgans (cM) in length with an average ratio of physical size to genetic map distance of 2.4 kb/cM. The *LEU2* locus is 45 ± 5 and 95 ± 5 cM from the short and long arm telomeres, respectively. R. K. Mortimer and D. Schild, *Microbiol. Rev.* **49**, 181 (1985); G. F. Carle and M. V. Olson, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3756 (1985).
 27. G. B. Dreyer and P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 968 (1985).
 28. The cleavage reaction conditions were as follows: yeast chromosomal DNA embedded in 0.5% low melting point agarose (~30- μ l blocks) was washed extensively with doubly distilled water to remove exogenous EDTA and equilibrated for 12 hours in 100 mM NaCl, 25 mM tris-acetate, pH 7.2 to 7.8, and 1.0 mM spermine-(HCl)₄ (NTS buffer). The agarose blocks were transferred to 2.0-ml flat-bottom eppendorf tubes and overlaid with 120 μ l of NTS buffer; 1.5 μ l each of 100 μ M oligonucleotide-(EDTA)₂ and 200 μ M Fe(NH₄)₂(SO₄)₂·6H₂O (freshly prepared) were pre-equilibrated for 5 min, added to the solution surrounding the agarose block, and incubated for 3 hours. After diffusion and triple-helix formation, the reaction was initiated by addition of 3 μ l of 200 mM DTT and allowed to proceed for 8 hours. The inactivated cleavage reagents were removed by three 30-min washes with 1-ml aliquots of 100 mM tris-HCl (pH 8.5), after which the agarose block was re-equilibrated in NTS buffer and a second reaction cycle performed with fresh oligonucleotide-(EDTA)₂ and DTT. All manipulations were performed at room temperature (22°C). After two cleavage cycles, the chromosomal blocks were washed in 1 ml of 10 mM tris-HCl (pH 7.5), 1 mM EDTA, and loaded on a 1% agarose (0.5 \times TBE) (19) gel for characterization of the cleavage products.
 29. J. K. Keeseey, R. Bigelis, G. R. Fink, *J. Biol. Chem.* **254**, 7427 (1979); R. J. Deshaies and R. Schekman, *J. Cell Biol.* **105**, 633 (1987).
 30. M. J. Dobson *et al.*, *Nucleic Acids Res.* **10**, 2625 (1982); M. F. Tuite *et al.*, *EMBO J.* **1**, 603 (1982).
 31. Secondary cleavage bands are more intense in the control (- target site) yeast strain because of the higher chromosomal DNA concentration in the sample. Thus band intensities can be compared within, but not between, yeast strains.
 32. We thank S. Emr for yeast strain SEY6210, the shuttle vector YEp13, and assistance in the design of the yeast transformation; R. Schekman for the *HIS4* marker in YCp503; J. Campbell for the *PGK1* promoter in pMA91; B. Birren for yeast strain YNN295; and assistance with pulsed-field gel electrophoresis; and the Howard Hughes Institute for a predoctoral fellowship to S.A.S. Supported by NIH grant GM 42966.

24 January 1990; accepted 1 May 1990

Comments to Site-Specific Cleavage of a Yeast Chromosome by Oligonucleotide-Directed Triple Helix Formation. Although a number of endogenous triple helix target sites could be identified within the yeast genome, including at least one within the *LEU2* gene itself (12), it was advantageous to engineer target sites into yeast chromosome III for a number of reasons: (i) a site of optimal length and A vs. G content could be targeted for initial studies on the larger DNA substrate, (ii) it provided a positive and a negative control for specific cleavage of a large DNA substrate that had not been fully mapped or sequenced, (iii) it enabled construction of chromosomes containing target sites compatible with triplex mediated enzymatic (15), alkylation (17, 18), and affinity cleaving techniques for cutting (1, 14, 19), and (iv) yeast constructs containing specific target sites from human DNA sequences were used to optimize conditions for cleaving the human genome (20).

The yields from affinity cleaving of yeast were significantly lower than observed with bacteriophage λ (1). One possible explanation stems from the use of high EDTA concentrations (0.5 M) in preparing the yeast chromosomes (21, 22). Unless all the EDTA was removed, the exogenous EDTA could compete for the Fe^{2+} , reducing the concentration of oligonucleotide-EDTA•Fe(II) complexes. A second possibility is that the bacteriophage λ yields were uncharacteristically high for this type of reaction. No other substrate has been cut with efficiencies comparable to λ phage. Plasmid DNA containing the yeast target site was cut at less than 15% cleavage, suggesting that the low efficiency was not inherent to the cleavage of megabase genomic DNA. A third possibility is that the target site was not saturated, however subsequent enzyme inhibition experiments proceeded almost quantitatively, arguing strongly against this possibility (15). Modest yields are more likely due to a

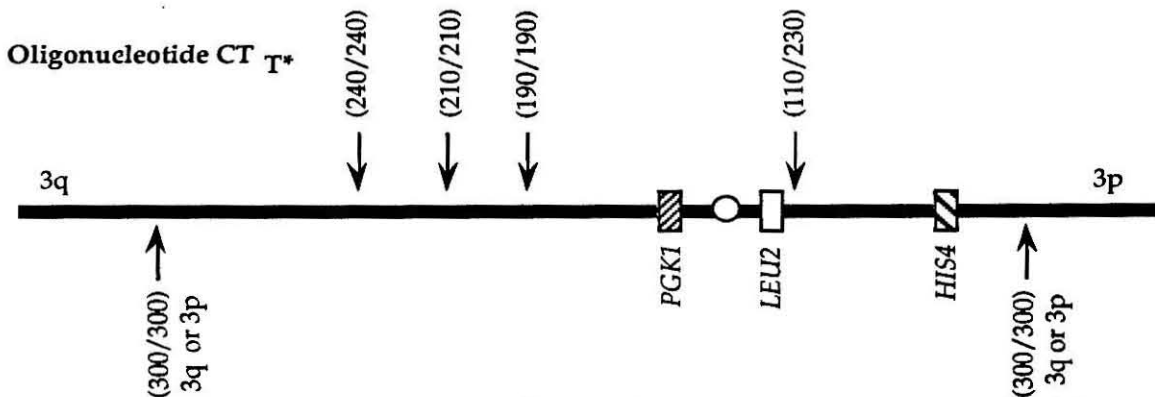


Fig. 3.1. Map of primary and secondary cleavage sites recognized by oligonucleotide CT_{T^*} on chromosome III at pH 7.0. The centromere (circle), *PGK1*, *LEU2*, and *HIS4* loci are shown as rectangles along chromosome III (dark line). The sites shown with arrows above the chromosome have defined map locations, while the site below the line could map to either 3p or 3q (both possible sites are shown). The pair of numbers next to each arrow correspond to the product sizes observed with *HIS4* and *PGK1* hybridizations, respectively.

combination of a target site less susceptible to triple helix mediated affinity cleaving and exogenous EDTA in the reaction competing for the metal cofactor.

In addition to the primary target site near the *LEU2* gene, the CT_{T^*} oligonucleotide was reported to cut at four secondary sites on chromosome III. Hybridization of the reaction products with *HIS4* (23, 24) and *PGK1* (25, 26) identified the location of the secondary cleavage sites (Fig. 3.1). The 300 kb product was not flanked by either of the markers and could not be definitively mapped to either end of chromosome III.

Cleavage of Yeast Chromosome III with MeC and BrU Modified Oligonucleotides. Oligonucleotides CT_{T^*} , $MeCT_{T^*}$ and $MeC^{BrU}T_{T^*}$ were synthesized specifically for the target site on yeast chromosome III with T^* at the 5' and 3' ends and MeC and BrU base modifications (27). Reactions were performed as a function of pH between 7.0 and 7.6 to examine the relative

specificity of the oligonucleotides (Fig. 3.2 A). Ethidium bromide staining of the cleavage products showed little chromosomal degradation by CT_T^* at any pH. $MeCT_T^*$ showed significant degradation of all chromosomes at pH 7.0, and modest degradation at pH 7.2 and above. $MeC^{Br}U_T^*$ extensively degraded the DNA, leaving no visibly intact chromosomal bands except at pH 7.6, where secondary cleavage remained extensive.

DNA hybridization with the *LEU2* (12, 28, 29) gene confirmed the general observations evident by ethidium bromide (Fig. 3.2 B). The oligonucleotide with the greatest specificity under the conditions tested was CT_T^* . At pH 7.6 only the 110 and 230 kb products were detected in significant yield (Strain SEY6210B (30) has a duplication of the *LEU2* locus that flanks the engineered target site resulting in detection of both the 230 and 110 kb cleavage products upon hybridization with *LEU2*). Oligonucleotide- $MeCT_T^*$ showed slightly less specificity, and generally higher overall background, particularly at lower pH. The specificity was partially restored at pH 7.6, though the secondary cleavage sites were cut more extensively than with CT_T^* at the same pH. $MeC^{Br}U_T^*$ was very nonspecific under these reaction conditions resulting in a broad smear in which all the chromosome III molecules had been cut at least once at pH 7.0. At pH 7.6, the specificity slightly improved, and specific bands became visible above the high background.

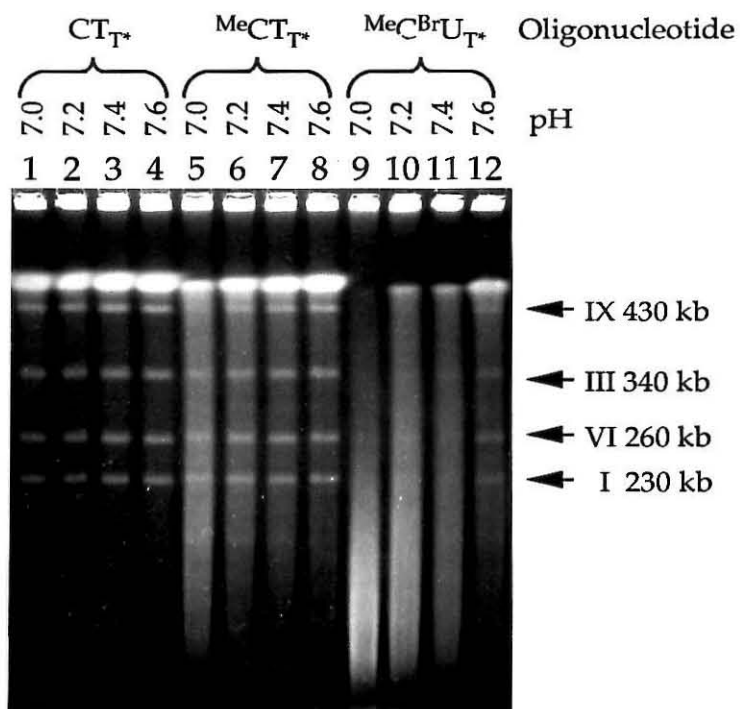
The relative specificities of the oligonucleotides was similar to that observed with bacteriophage λ cleavage (1) (Chapter 2). It suggests that both pH and oligonucleotide composition can be used to adjust the specificity of

Fig. 3.2 Triple helix mediated affinity cleaving of yeast chromosomal DNA prepared from strain SEY6210B as a function of oligonucleotide composition and pH.

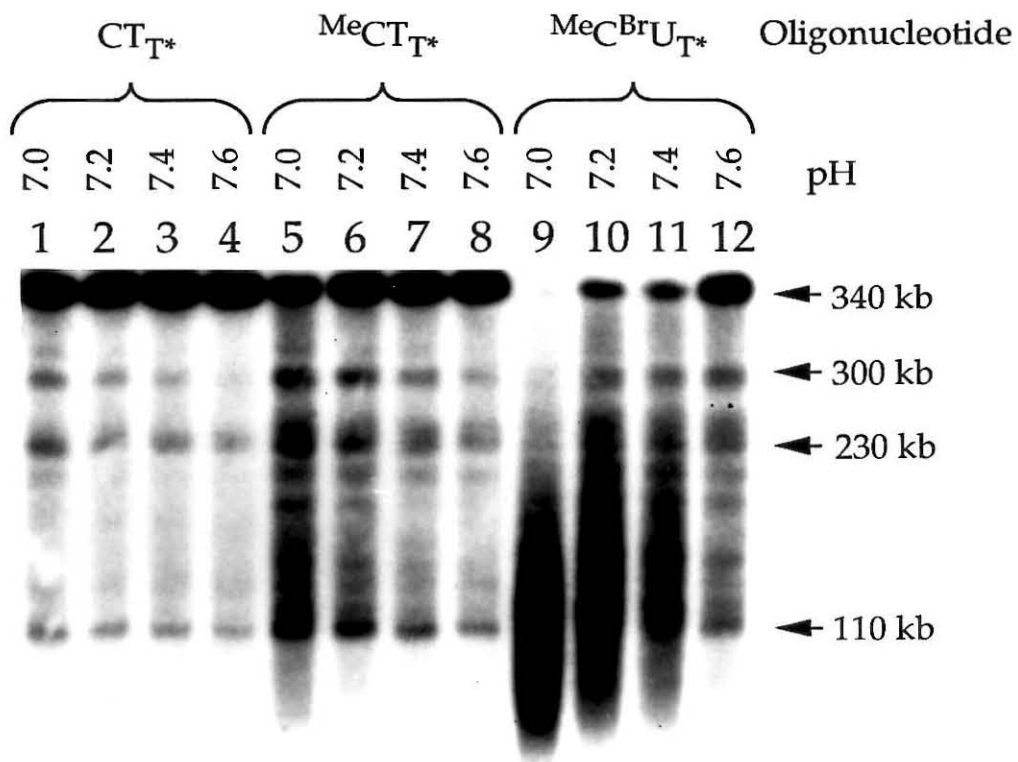
A. Ethidium bromide stained gel of total yeast chromosomal DNA reacted with oligonucleotides CT_T^* , $MeCT_T^*$, and $MeCBrU_T^*$ at pH 7.0 to 7.6 (conditions in each lane listed in legend above gel). Chromosomal DNA embedded in 0.6% Incert LMP agarose was washed extensively to remove exogenous EDTA and equilibrated for 12 hours in 100 mM NaCl, 25 mM tris-acetate, pH 7.0 to 7.6, and 1.0 mM spermine•4HCl (NTS buffer). The agarose blocks ($\approx 80 \mu\text{l}$) were transferred to 2.0-ml flat bottom eppendorf tubes and overlaid with 120 μl of NTS buffer. 2.5 μl of 100 μM oligonucleotide (CT_T^* , $MeCT_T^*$, or $MeCBrU_T^*$) was equilibrated with 2.5 μl of 100 μM $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ (freshly prepared) for 5 minutes, and 4 μl of the resulting solution added to the DNA. The solution was incubated for 3 hours to promote triple helix formation, and the affinity cleaving reaction was initiated by the addition of 4 μl of 200 mM DTT (freshly prepared). The cleavage reaction was allowed to proceed for 8 hours. The DNA was digested a second time with fresh reagents by washing agarose plugs with 1.0 ml aliquots of 100 mM tris-HCl (pH 8.5) for 30 min. three times, and reequilibrating the plugs in NTS buffer. The plugs were then incubated with a second aliquot of oligonucleotide, $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ and DTT. Reactions were performed at room temperature (22° C). After two reaction cycles, the chromosomal plugs were washed in 1 ml of 10 mM tris-HCl (pH 7.5), 1 mM EDTA, and loaded on a 1% agarose (0.5xTBE) gel at 6.0 V/cm, 120° switch angle, 15-30 sec. ramped switch times for 25 hours at 14° C. The DNA stained with ethidium bromide and visualized by UV irradiation.

B. DNA hybridization of gel shown in A, hybridized with the PCR product of the *LEU2* subcloned plasmid pUCKE400 (see methods for plasmid construction and hybridization protocol). A duplication of the *LEU2* gene in construct SEY6210B resulted in the detection of both the 110 and 230 kb products upon hybridization with *LEU2*. Other secondary cleavage sites are also evident as a function of pH and oligonucleotide composition.

A



B



triple helix formation. This observation was useful for the identification of endogenous target sites in unsequenced human DNA (20) (Chapter 4).

Cleavage of Yeast Chromosomal DNA with Nonspecific Oligonucleotides. The observation that an oligonucleotide was capable of cleaving sites for which it was not specifically designed and not completely homologous (14), suggested that it should be possible to use nonspecific oligonucleotides to cut yeast chromosomal DNA. To test this, oligonucleotides specific for sites in bacteriophage λ (oligonucleotides 1 and 2) and adenovirus (oligonucleotides 3 and 4) genomes were used to cut yeast chromosome III (Fig. 3. 3A). No degradation was apparent by ethidium bromide staining for reactions with oligonucleotides 1, 2, and 3. Oligonucleotide 4, containing a long T sequence, produced a faint smear consistent with a large number inefficiently cut sites. This oligonucleotide is specific for poly-A runs frequently found in genomic DNA.

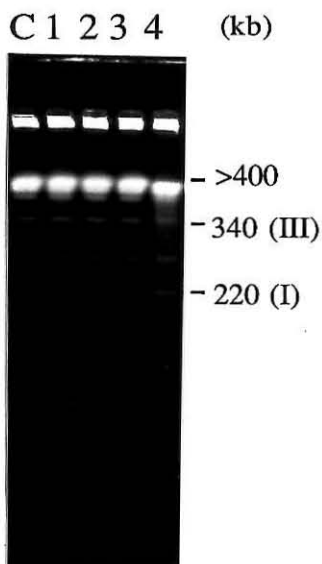
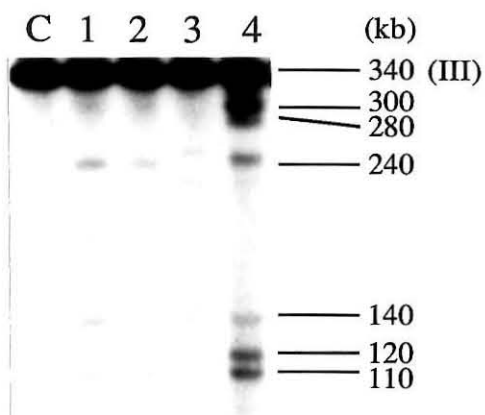
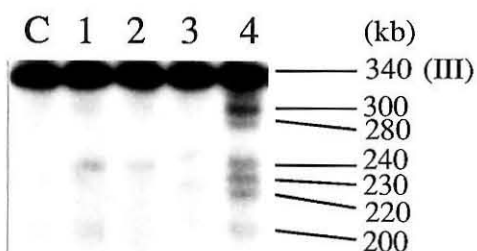
DNA hybridization with chromosome III specific markers *PGK1* and *HIS4* revealed a discrete set of cleavage products generated by the oligonucleotides (Fig. 3.3 B). Oligonucleotides 1 and 2, differing only in the number of T* moieties, produced identical products with different efficiencies. Oligonucleotide 2 cut less efficiently, suggesting that the second T* was destabilizing for these target sites. Oligonucleotide 3, containing a high C content, inefficiently cut only two sites suggesting low oligonucleotide affinity and minimal sequence similarity. In contrast, the poly-T oligonucleotide 4 produced six discrete cleavage products in good yield.

Comparison of *PGK1* and *HIS4* hybridizations allowed most of the sites to be mapped to a specific position on chromosome III (Fig. 3.4). Oligonucleotides 1, 2 and 4 bound a sequence at or near the engineered target site producing 110 (*HIS4*) and 230 (*PGK1*) kb cleavage products. It is

Fig. 3.3. Cleavage of yeast chromosomal DNA with nonspecific oligonucleotides by affinity cleaving. **Top.** Oligonucleotide-EDTA sequences specific for target sites in other substrates. Oligonucleotides **1** and **2** were designed to cut bacteriophage λ DNA at the origin of replication. These oligonucleotides were originally given the title λ_5' and $\lambda_5'3'$, respectively. Oligonucleotides **3** and **4** were designed for target sites in adenovirus-2 genomic DNA. They were originally titled Ad-1 and Ad-2, respectively. **Middle.** Ethidium bromide staining of cleavage reactions using oligonucleotides **1-4** (indicated above gel) to cut chromosomal DNA prepared from SEY6210B. Cleavage reactions and pulsed-field gel electrophoresis were performed as described in Fig. 3.2 except only one cleavage cycle was completed. **Bottom.** DNA hybridization of gel shown above with DNA markers HIS4 and PGK1. DNA transfer, immobilization, and hybridization were performed as described in the methods section of this chapter. Cleavage products were detected by autoradiography, and size estimates made by comparison to a bacteriophage λ DNA ladder.

Oligonucleotides

1. 5'- T*TTTCTTTTTTCTTTTCT-3'
2. 5'- T*TTTCTTTTTTCTTTTCT* -3'
3. 5'- T*CTTCTTCTTCTTCTTCCCCTCC-3'
4. 5'- T*TTTTTTTTTTTTTTTC-3'

*HIS4**PGK1*

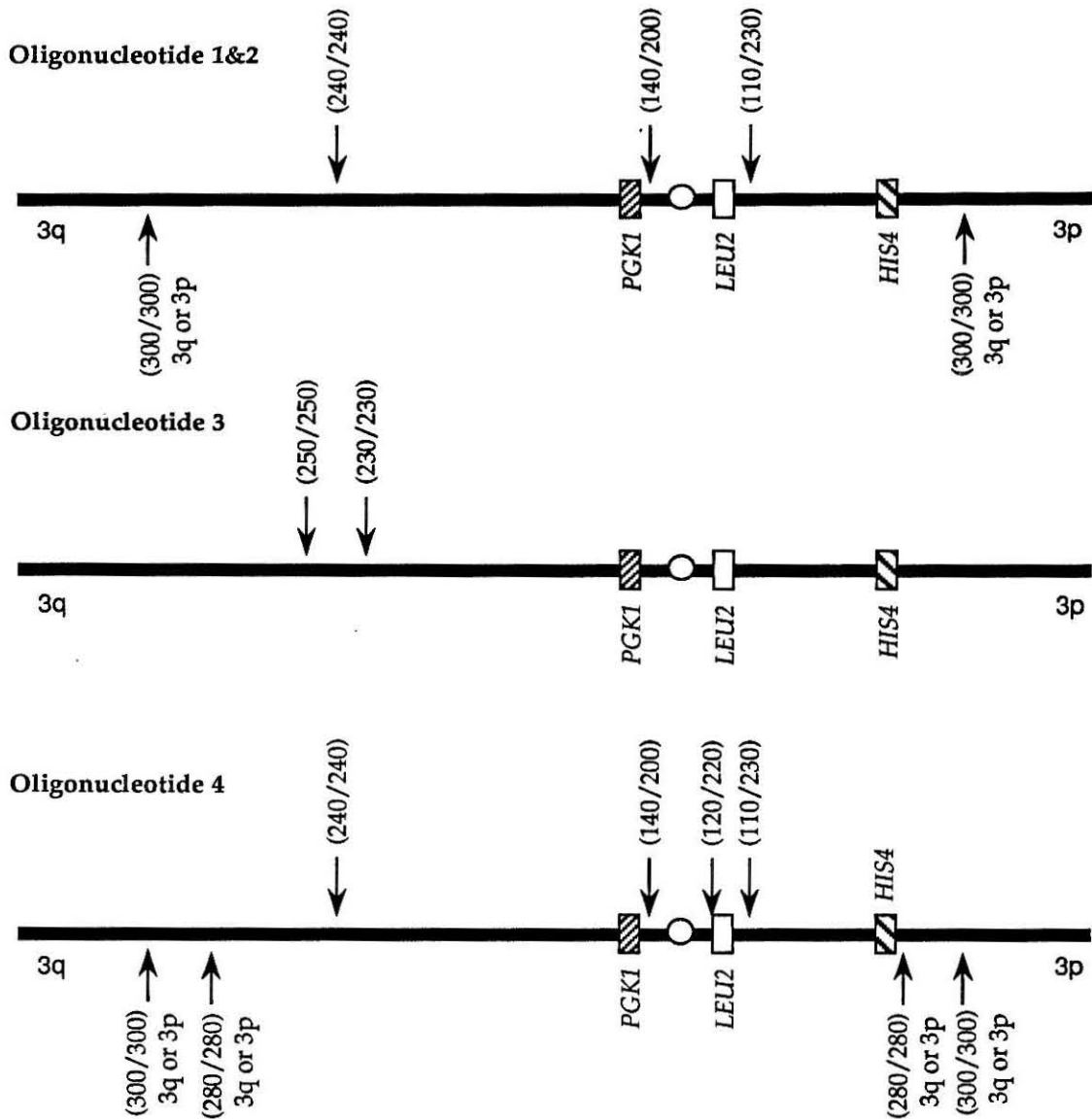


Fig. 3.4. Map of primary and secondary cleavage sites recognized by oligonucleotides 1-4 on chromosome III at pH 7.0. The centromere (circle), *PGK1*, *LEU2*, and *HIS4* (rectangles) loci are shown along chromosome III (dark line). The sites shown with arrows above the chromosome have defined map locations, while sites below the line could map to either 3p or 3q (both possible sites are shown). The pair of numbers next to each arrow correspond to the product sizes observed with *HIS4* and *PGK1* hybridizations, respectively.

noteworthy that there is a long poly-A run about 300 bp 5' of the target site that could be recognized by the oligonucleotides (12). All of the sites recognized by oligonucleotides 1 and 2 were cut more efficiently by oligonucleotide 4, suggesting that the sites bound were predominated by A rich sequences. This is indirectly supported by the observation that sites recognized by oligonucleotide 3, a sequence with high-C content, were not detected by any of the other oligonucleotides.

The cleavage pattern demonstrates that T* oligonucleotides are useful for generating a partial DNA digest whose products depend upon the sequence of the oligonucleotide, without requiring prior knowledge of sequence information. Affinity cleaving mediated by nonspecific oligonucleotides is similar to restriction enzyme mapping by partial digestion. This could potentially be useful if affinity cleaving oligonucleotides designed for conserved sequence elements within protein coding regions could be used to map the location and frequency of such sequences within the genome.

Conclusions. This work demonstrated the feasibility of targeting specific sites within megabase genomic DNA by triple helix formation with small oligonucleotides. Oligonucleotide binding was assayed by affinity cleaving of yeast chromosome III. In addition to the target site, affinity cleaving detected four secondary binding sites on yeast chromosome III and DNA hybridization facilitated the identification of their map position. The specificity of oligonucleotide directed triple helix formation could be regulated by adjusting the pH and the base composition of the oligonucleotides. Although the low efficiency of cutting by this method makes generating large quantities of cleaved genomic material impractical, affinity cleaving demonstrated the specificity of triple helix formation on

large DNA, and could be useful for mapping the frequency and location of conserved sequence elements within chromosomes.

Part II

Single Site Enzymatic Cleavage of Yeast Genomic DNA Mediated by Triple Helix Formation

While affinity cleaving was useful for demonstrating triple helix formation at specific sites within large genomic DNA, it was limited by low cleavage efficiency and cutting at secondary sites (1, 14). Because affinity cleaving cut all sites bound by the T* end of the oligonucleotide, it demonstrated the gap between the theoretical frequency of the target site (one site in 10^{11} bp) (31) compared to the actual ability of the oligonucleotide to differentiate between several similar but imperfectly matched binding sites (approximately one site in 10^5 bp). Thus, despite the size of the recognition site, affinity cleaving mediated by oligonucleotide directed triple helix formation was insufficient to define a single site in megabase DNA.

Single site cleavage of megabase DNA was achieved by two techniques that coupled the specificity of triple helix formation with a sequence specific cleaving function (15-18). Triple helix mediated alkylation combined triple helix formation with the requirement for a G (alkylation at N7) two bases to the 5' end of the oligonucleotide (Appendix B) (17). Work done in collaboration with Povsic demonstrated that double strand cleavage was only observed if the sequence met the following criterion: (i) there must be two oligonucleotide binding sites, (ii) the oligonucleotide binding sites must be on opposite strands, (iii) there must be a G two bp to the 5' side of each binding site, and (iv) the binding sites must be within approximately 20 bp of each

other to facilitate strand separation (32). Such a rigidly defined class of sequences occurs sufficiently infrequently to specify a single site in megabase DNA. Alkylation was highly efficient, producing double strand cleavage yields approaching 90% (Appendix B).

A second strategy for generating quantitative single site cleavage combined triple helix formation with the specificity of methylases and restriction enzymes (15, 33, 34). In this approach, a triple helix site partially overlapping the recognition site of a sequence specific methylase blocked the methylase activity only at the site bound by the oligonucleotide. All other methylation sites were modified by the enzyme. Subsequent methylase inactivation, triple helix disruption, and digestion by the cognate restriction enzyme produced cleavage only at the unmodified site previously bound by the oligonucleotide. In this approach, cleavage is only observed if the oligonucleotide binding site partially overlaps the methylation site. If there is no overlap, or if the site bound is not adjacent to a methylation site, no cleavage is observed. This section describes the application of this technique to the cleavage of yeast chromosome III at a single site.

Site Specific Cleavage of Plasmid DNA Containing a Triple Helix Target Site. Koob and Szybalski originally demonstrated that protein mediated methylase inhibition could be used to modify the effective specificity of a restriction enzyme on plasmid and yeast genomic substrates (15, 33, 34). They entitled the procedure Achilles' heel cleavage because of its similarity to the story of Achilles from Greek mythology. Maher, Wold, and Dervan demonstrated the feasibility of triple helix mediated methylase inhibition on plasmid DNA using the cognate enzymes TaqI methylase and TaqI endonuclease (35, 36). Although this conceptually verified that triple helix mediated Achilles' heel cleavage was also possible, it utilized the unique thermostability of the restriction enzyme to remove the oligonucleotide. Therefore, it was necessary to optimize triple helix and enzyme conditions for thermolabile enzyme pairs. The temperatures used to destabilize the triplex were incompatible with large DNA as the LMP agarose used to retain chromosomal integrity would melt at 65° C. To optimize low temperature reaction conditions, the plasmid used for homologous recombination with chromosome III, pUCLEU2A, was used as a cleavage reaction substrate.

Plasmid pUCLEU2A was generated by inserting a homopurine sequence overlapping an EcoRI methylase/endonuclease site into the unique XhoI site of pUCLEU2 (Fig. 3.5) (see methods). The plasmid contained a total of three EcoRI sites, as well as a unique HindIII site immediately opposite the EcoRI target site. Thus, if the DNA was linearized with HindIII and cut only at the desired EcoRI site, the linear 6.6 kb plasmid would be cut into two 3.3 kb products. The ratio of the 6.6 to the 3.3 kb band detected by ethidium bromide staining would define the cleavage efficiency.

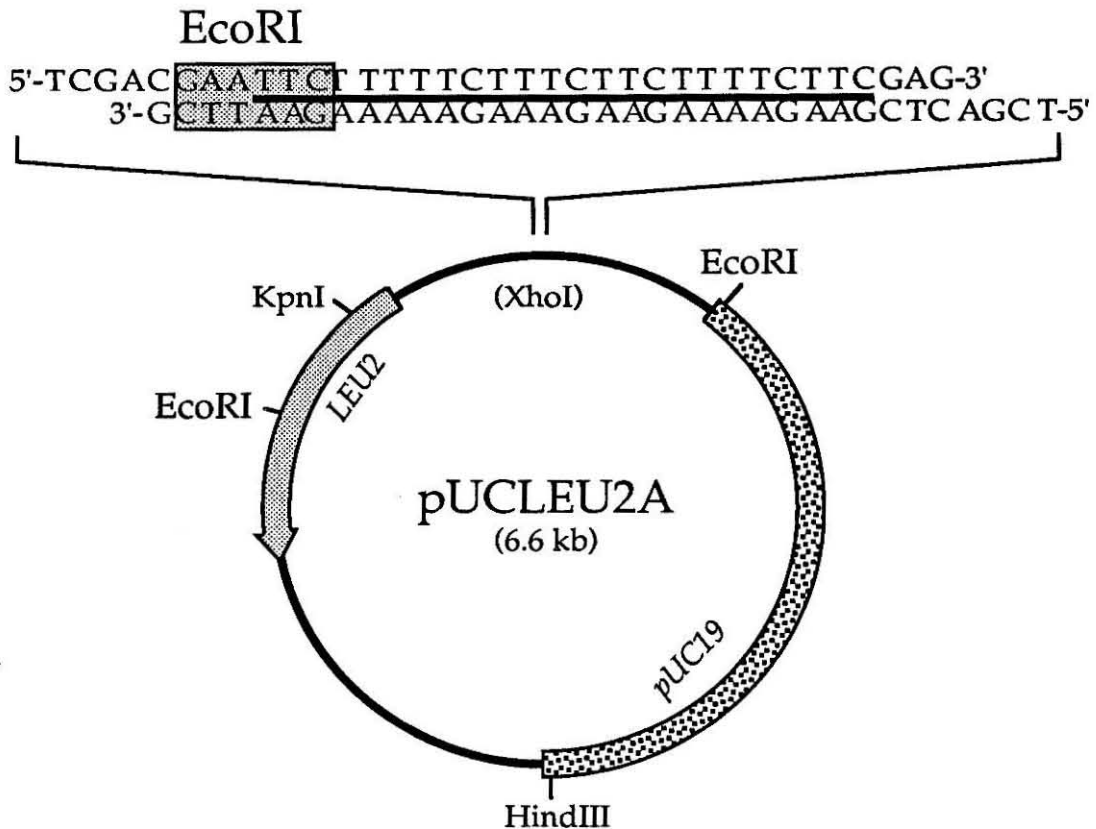


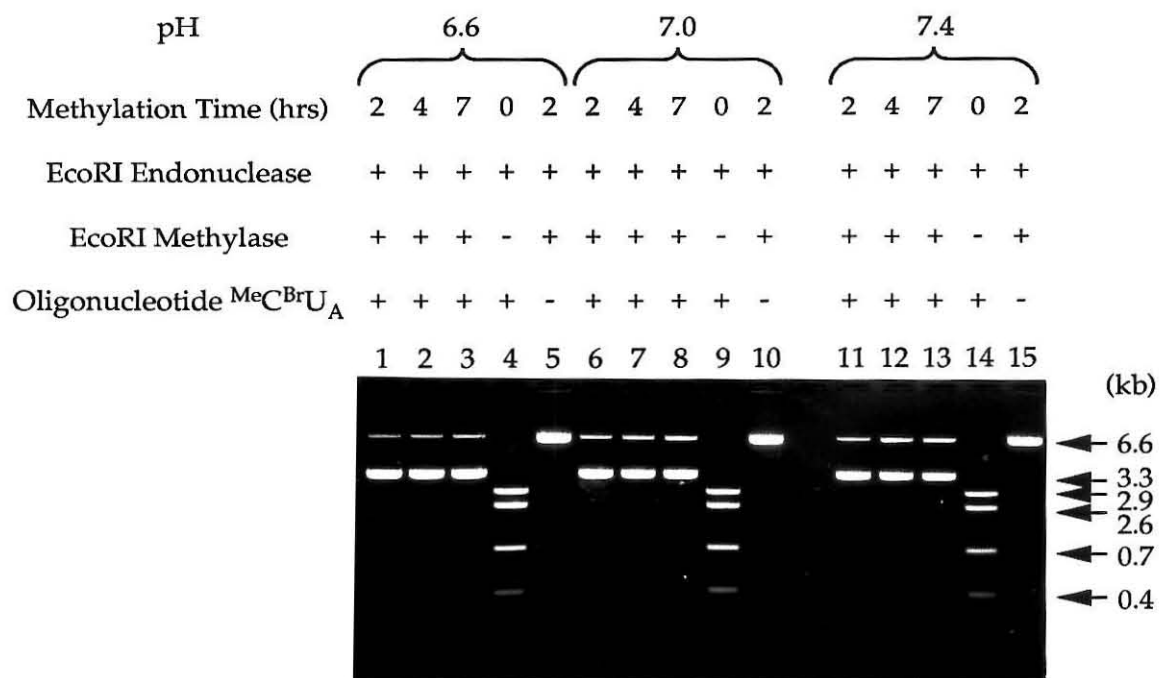
Fig. 3.5. Map of pUCLEU2A showing the triple helix target site. The target EcoRI site is shown in a shaded box within the duplex sequence inserted into the XhoI site (top). The triple helix binding site is shown as a dark line within the duplex. Also shown are the restriction sites used in this study, the *LEU2* gene, the pUC19 vector, and the size of the plasmid.

It was initially important to define a strategy for the removal of the oligonucleotide from the target site. LMP agarose has an upper temperature limit of 55° C, a temperature too low for complete triple helix disruption. Triple helix formation was shown to be polycation sensitive suggesting that removal of the spermine from the solution should be sufficient to destabilize the triplex (19). This approach was problematic because the restriction enzyme buffer contains magnesium, a dication capable of triplex stabilization. If oligonucleotide was still present upon addition of endonuclease buffer, the oligonucleotide could reform the triplex, block the restriction enzyme, and

reduce the cleavage efficiency. Several studies demonstrated that triple helix formation is highly pH sensitive (19, 27, 37). By raising the pH of the restriction enzyme digestion buffer to between 8.5 - 9.5, the oligonucleotide was removed from the target site, without effecting the fidelity of the restriction enzyme.

Reactions performed as a function of methylation time and pH are shown in Fig. 3.6. No cleavage was observed in the absence of oligonucleotide indicating that the methylase had reacted to completion (lanes 5, 10, 15). Complete digestion of plasmid DNA was observed in the absence of methylase despite the addition of oligonucleotide (lanes 4, 9, 14). This indicates that the high pH workup conditions were sufficient to remove the oligonucleotide from the duplex target site, that the specificity of the restriction enzyme was not significantly affected by the high pH buffer, and that EcoRI was cutting the DNA to completion. Plasmid DNA methylated in the presence of the oligonucleotide was cut by EcoRI endonuclease only at the triplex target site to generate the desired 3.3 kb products in $92 \pm 3\%$ yield (lanes 1-3, 6-8, and 11-13). After longer methylation times, the cleavage efficiency was gradually reduced, suggestive of a slow oligonucleotide off rate (38) (compare lanes 1 & 3, 6 & 8, and 11 & 13). The cleavage efficiency was slightly better at pH 6.6 than at pH 7.4 indicative of triple helix stabilization at low pH. However, at lower pH the methylase required a slightly longer incubation time to achieve complete methylation (lane 1). While the techniques used at the plasmid level could not be immediately transferred to yeast chromosomal DNA, i.e., the reactions were phenol extracted and ethanol precipitated, two procedures incompatible with agarose embedded chromosomal DNA, they demonstrated that an oligonucleotide was capable of site specific protection

Fig. 3.6. Enzymatic cleavage of plasmid pUCLEU2A DNA mediated by triple helix formation. Reagents, methylation times, and pH were as indicated above the figure. pUCLEU2A plasmid DNA (2 μg) was incubated for 2.5 hours with 1 μM MeCBrU_A oligonucleotide in methylation buffer containing 100 mM NaCl, 100 mM tris-HCl, 100 $\mu\text{g}/\text{ml}$ BSA, 2 mM spermine, 10 mM EDTA, pH 6.6, 7.0, or 7.4 at 22° C. Following triple helix formation 160 μM S-adenosyl methionine (SAM) and 120 units of EcoRI methylase were added resulting in a final volume of 20 μl . The DNA was methylated for 2, 4, or 7 hours at 22° C after which the methylase was stopped by phenol extraction. The DNA was recovered by ethanol precipitation, 70% ethanol washing, and drying. The DNA was resuspended in EcoRI restriction enzyme buffer containing 10 units EcoRI, 100 $\mu\text{g}/\text{ml}$ BSA, 100 mM tris-HCl, pH 9.5, 50 mM NaCl, 5 mM MgCl₂ and digested to completion at 37° C for 45 min. The products were loaded directly on a 0.8% agarose gel containing 1xTAE and ethidium bromide. Products were visualized by UV illumination.



against *EcoRI* methylation resulting in highly efficient cleavage upon *EcoRI* restriction enzyme digestion.

Comparable experiments were performed on plasmid DNA using enzymes *HpaII*, *HaeIII*, *MspI*, *dam/MboI*, *AluI*, *TaqI*, *AluI-Me/SstI*, and *HindIII*. Nonoptimized cleavage efficiencies ranging from 50-98% were observed. In each system it was necessary to reduce the pH of the manufacture's recommended buffer to near pH 7.0, add spermine to promote triple helix formation, and add at least 50 mM NaCl to prevent DNA precipitation by the spermine.

Single-site enzymatic cleavage of yeast genomic DNA mediated by triple helix formation

Scott A. Strobel & Peter B. Dervan*

Division of Chemistry and Chemical Engineering,
California Institute of Technology, Pasadena,
California 91125, USA

PHYSICAL mapping of chromosomes would 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 generalizability of the recognition motif, and the cleavage yield. Oligonucleotide-directed triple helix formation is a generalizable motif for specific binding to sequences longer than 12 base pairs within DNA of high complexity¹⁻³. Studies with plasmid DNA show that triple helix formation can limit the operational specificity of restriction enzymes to endonuclease recognition sequences that overlap oligonucleotide-binding sites^{4,5}. Triple helix formation, followed by methylase protection, triple helix-disruption, and restriction endonuclease digestion produces near quantitative cleavage at the single overlapping triple helix-endonuclease site^{4,5}. As a demonstration that this technique may be applicable to the orchestrated cleavage of large genomic DNA, we report the near quantitative single-site enzymatic cleavage of the *Saccharomyces cerevisiae* genome mediated by triple helix formation. The 340-kilobase yeast chromosome III was cut uniquely at an overlapping homopurine-*EcoRI* target site 27 base pairs long to produce two expected cleavage products of 110 and 230 kilobases. No cleavage of any other chromosome was detected. The potential generalizability of this technique, which is capable of near quantitative cleavage at a single site in at least 14 megabase pairs of DNA, could enable selected regions of chromosomal DNA to be isolated without extensive screening of genomic libraries.

Chemical reagents and biological methods that infrequently cleave double helical DNA are being developed for genomic mapping¹⁻¹². Recently, Szybalski and coworkers reported an 'Achilles heel cleavage' technique that limits restriction enzyme cleavage to an overlapping *lac* repressor-endonuclease site^{11,12}. Protein-mediated Achilles heel cleavage may, however, not be readily generalized to unique cleavage at selected genetic markers owing to the paucity of applicable DNA-binding proteins relative to the number of sites in megabase genomic DNA. To achieve orchestrated cleavage at selected genetic markers in human DNA by this approach may require the protein-binding sequence to be artificially inserted into the chromosome at precise locations. A more general method for recognition of endogenous DNA sequences might be a chemical approach based on oligonucleotide-directed triple helix formation¹. Pyrimidine oligonucleotides bind specifically to purine sequences in duplex DNA to form a local triple-helix structure^{1-5,13,14}. The generalizable code for triple-helix specificity is derived from thymine (T) binding to adenine-thymine base pairs (T·AT base triplet)¹⁵ and N3 protonated cytosine (C+) binding to guanine-cytosine base pairs (C+GC base triplet)^{16,17}. Higher affinity oligonucleotides can be obtained by substituting 5-bromouracil for thymine and 5-methylcytosine for cytosine¹⁸. Moreover, the number of potential target sequences amenable to recognition by the triple helix motif can be extended to some mixed purine-pyrimidine sequences^{19,20}. By replacing the *lac* repressor protein in the Achilles heel technique with triple-helix-mediated protection, one combines the strengths of the two approaches; a general chemical method for recognition of DNA sequences in the range of 15-20 base pairs (bp) with the cleavage efficiency of restriction enzymes (Fig. 1).

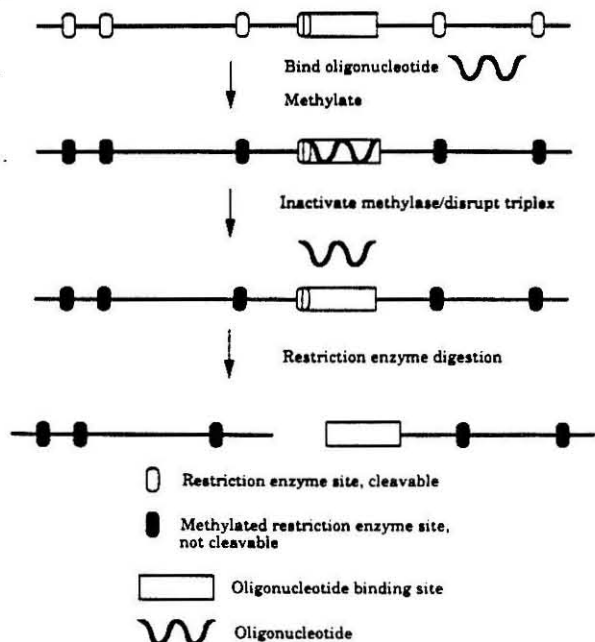


FIG. 1 General scheme for single-site enzymatic cleavage of genomic DNA by oligonucleotide-directed triple helix formation. Chromosomal DNA is equilibrated with an oligonucleotide in a methylase compatible buffer containing polycation. *EcoRI* methylase which methylates the central adenines of the sequence 5'-GAATTC-3' and renders the sequence resistant to cleavage by *EcoRI* 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 *EcoRI* restriction endonuclease. The cleavage products are separated by pulsed-field gel electrophoresis and efficiencies quantitated by Southern blotting.

To demonstrate single-site enzymatic cleavage of the yeast genome by this technique, a sequence containing an overlapping 24-bp purine tract and 6-bp *EcoRI* site was inserted proximal to the *LEU2* (ref. 21) gene on the short arm of chromosome III (Fig. 2) (ref. 3). Oligonucleotides designed to form triple helix complexes that overlap half of the *EcoRI* recognition site were synthesized with CT, ^mC^T or ^mC^{Br}U nucleotides¹⁸. The genetic map of yeast chromosome III (refs 22, 23) and affinity cleaving data³ indicate that cleavage at the target site should produce two fragments 110 ± 10 and 230 ± 10 kilobases (kb) in size (Fig. 2).

Resolution of total yeast chromosomal DNA by pulsed-field gel electrophoresis^{24,25} revealed that chromosome III was cut exclusively at the target site when a ^mC^T oligonucleotide was used for triple helix formation at pH 7.6 (Fig. 3, lane 1). No cleavage was detected on any other chromosomes under these conditions nor was cleavage observed in the absence of oligonucleotide (lane 2) or in a yeast strain lacking the target sequence (lane 3). The expected 110-kb product was visualized with ethidium bromide staining and confirmed by Southern blotting with a *HIS4* (ref. 26) marker (Fig. 3c). The 230-kb product comigrated with chromosome I, but was detected by Southern blotting with a *LEU2* (ref. 24) marker (Fig. 3b). The cleavage efficiency was 94 ± 2%. Similar efficiencies were seen with a CT oligonucleotide up to pH 7.4 and ^mC^T and ^mC^{Br}U oligonucleotides past pH 7.8 (Fig. 4). The cleavage efficiency with all oligonucleotides was gradually reduced with longer methylation times, suggesting that the oligonucleotide dissociation rate might be the limiting factor for efficiency in this system²⁷.

The specificity of triple helix formation has been shown to be pH-dependent¹⁻³. By lowering the pH, sequences of near but imperfect similarity can be bound and cleaved¹⁻³. In agreement

* To whom correspondence should be addressed.

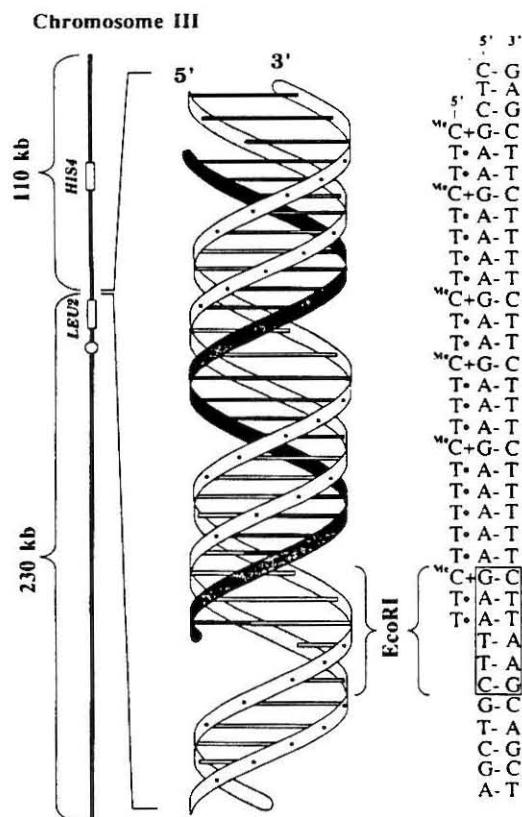


FIG. 2 Left, genetic map of *S. cerevisiae* chromosome III. The locations of the *HIS4* and *LEU2* loci (boxes), the centromere (circle), and the triple helix-EcoRI target site are indicated. The expected sizes of the cleavage products are shown. Right, schematic diagram of the triple helix complex overlapping the *EcoRI* restriction-methylation site. The pyrimidine oligonucleotide is bound in the major groove parallel to the purine strand of the DNA duplex, and covers half the *EcoRI* site.

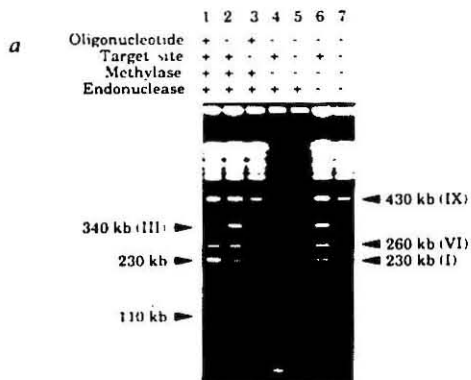
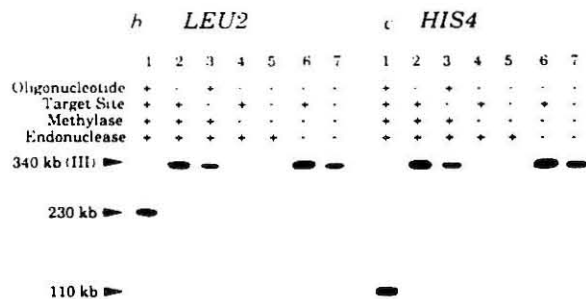


FIG. 3 Single-site enzymatic cleavage of yeast genomic DNA with reagents indicated above figure. 1. Yeast chromosomal DNA embedded in low-melting-point agarose ($\sim 50 \mu\text{l}$) was washed twice in 1.0 ml of 100 mM NaCl, 100 mM Tris-HCl, 10 mM EDTA, 2 mM spermine, pH 7.6 for 10 min, decanted, and overlaid with 150 μl of the same buffer. Oligonucleotide was added to 1 μM final concentration and incubated 8 h at 22 $^\circ\text{C}$. 2. Bovine serum albumin (100 $\mu\text{g ml}^{-1}$), *S*-adenosylmethionine (160 μM) and *EcoRI* methylase (80 units) were added to the overlay and incubated with shaking for 3 h. 3. The triple helix was destabilized and the methylase was simultaneously inactivated at 55 $^\circ\text{C}$ in 1.0 ml of 1% lauryl sarcosyl, 100 mM Tris-HCl pH 9.5, 10 mM EDTA for 30 min. The oligonucleotide and detergent were then removed with 4 \times 10 min washes (1.0 ml) with 10 mM Tris-HCl pH 9.5, 10 mM



EDTA. 4. The agarose plug was washed twice with 1.0 ml 2 \times potassium glutamate restriction enzyme buffer (200 mM potassium glutamate, 50 mM Tris-acetate, pH 7.6, 20 mM magnesium acetate, 100 $\mu\text{g ml}^{-1}$ BSA, 1 mM 2-mercaptoethanol), decanted, overlaid with 150 μl of buffer and the DNA digested to completion with 20 U *EcoRI* restriction endonuclease for 6 h at 22 $^\circ\text{C}$. After digestion the enzyme was heat-inactivated at 55 $^\circ\text{C}$ for 10 min and the DNA loaded onto a 1% agarose, 0.5 \times TBE pulsed-field gel at 14 $^\circ\text{C}$, 200 V. Switch times were ramped from 10–40 s for 18 h followed by ramping from 60–90 s for 6 h. Products were visualized by ethidium bromide staining (a) and Southern blotting with *LEU2* (b) and *HIS4* (c) chromosome markers. Cleavage efficiencies were measured using storage phosphor imaging plates with a Molecular Dynamics 400S PhosphorImager.

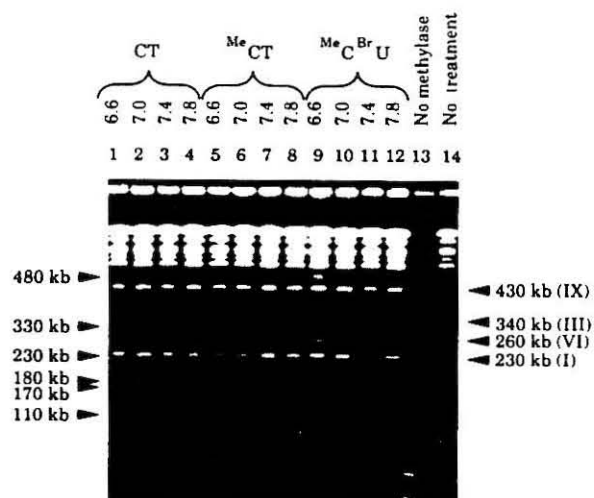


FIG. 4 Triple-helix-mediated enzymatic cleavage of the yeast genome as a function of oligonucleotide composition and pH. Lanes 1–12, reactions on a yeast strain containing the triple helix target site with oligonucleotides (CT, $^{\text{Me}}$ CT and $^{\text{Me}}$ C Br U) and pH values (6.6, 7.0, 7.4 and 7.8) as indicated above figure. Methylation time was 4.5 h. 170-kb and 650-kb (unresolved) secondary cleavage products were observed with both $^{\text{Me}}$ C substituted oligonucleotides at or below pH 7.4. The cleavage site can be assigned to chromosome II (820 kb) by two-dimensional pulsed-field gel electrophoresis in which the DNA was triple helix-protected and methylated before the first dimension and restriction enzyme-digested before the second dimension (data not shown). Additional secondary cleavage sites were observed with the $^{\text{Me}}$ C Br U oligonucleotide at pH 6.6 (lane 9). The 180- and 480-kb products were assigned to chromosome XI (660 kb), and the 330 and 780 kb (unresolved) products were assigned to the VII/XV doublet (1,100 kb) by the same method (data not shown).

with this observation, secondary cleavage sites were revealed as a function of oligonucleotide composition and pH (Fig. 4). Cleavage at the secondary site was preferentially reduced by longer methylation times, suggesting differential oligonucleotide dissociation rates between the primary and secondary target sites. Cleavage at all secondary sites could be eliminated at a threshold pH for each oligonucleotide.

Whereas affinity cleaving using oligonucleotide-EDTA·Fe identifies all sites of oligonucleotide binding¹⁻³, triple-helix-mediated endonuclease cleavage exposes only those sequences that also partially overlap a restriction site^{4,5}. The sequence requirement of a methylation-restriction site increases the cleavage specificity but reduces the number of available sites. To partially overcome this limitation, other commercially available methylation and restriction enzyme pairs with homopurine half sites were tested on plasmid DNA. In addition to *TaqI* (ref. 4) and *EcoRI* (ref. 5), single-site protection of plasmid DNA was possible with *MspI*, *HpaII* and *AluI* methylases.

BamHI, *HaeIII* and *dam* methylases could potentially be used, but remain untested.

The generalizability of 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. The use of degenerate oligonucleotides in this technique to rapidly screen 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 (S.A.S. and P.B.D., unpublished observations). □

Received 5 October; accepted 20 December 1990.

- Moser, H. E. & Dervan, P. B. *Science* **238**, 645-650 (1987).
- Strobel, S. A., Moser, H. E. & Dervan, P. B. *J. Am. chem. Soc.* **110**, 7927-7929 (1988).
- Strobel, S. A. & Dervan, P. B. *Science* **248**, 73-75 (1990).
- Maher, L. J., Wold, B. & Dervan, P. B. *Science* **248**, 725-730 (1989).
- Harvey, J. C., Shmizu, M. & Wells, R. D. *Nucleic Acids Res.* **18**, 157-161 (1990).
- Barlow, D. P. & Lehrach, H. *Trends Genet.* **3**, 167-171 (1987).
- Wenzlau, J. M., Seldanha, R. J., Butow, R. A. & Periman, P. S. *Cell* **66**, 421-430 (1989).
- Deishodde, A. *et al. Cell* **66**, 431-441 (1989).
- Patel, Y., Van Cott, E., Wilson, G. G. & McClelland, M. *Nucleic Acids Res.* **18**, 1603-1607 (1990).
- Weil, M. D. & McClelland, M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 51-55 (1989).
- Koob, M., Grimes, E. & Szybalski, W. *Science* **241**, 1064-1066 (1988).
- Koob, M. & Szybalski, W. *Science* **260**, 271-273 (1990).
- Praseuth, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 1349-1353 (1988).
- Lyamichov, V. I. *et al. Nucleic Acids Res.* **16**, 2155-2178 (1988).
- Felsenfeld, G., Davies, D. R. & Rich, A. *J. Am. chem. Soc.* **79**, 2023-2024 (1957).
- Lipsitt, M. N. *J. Biol. Chem.* **239**, 1256-1260 (1964).
- Rajagopal, P. & Feigon, J. *Nature* **338**, 637-640 (1989).
- Povsic, T. J. & Dervan, P. B. *J. Am. chem. Soc.* **111**, 3059-3061 (1989).
- Griffin, L. C. & Dervan, P. B. *Science* **248**, 967-971 (1989).
- Horne, D. A. & Dervan, P. B. *J. Am. chem. Soc.* **112**, 2435-2437 (1990).
- Andreadis, A., Hsu, Y. P., Kohlhaw, G. B. & Schimmel, P. *Cell* **31**, 319-325 (1982).
- Mortimer, R. K. & Schild, D. *Microbiol. Rev.* **48**, 181-212 (1985).
- Carle, G. F. & Olson, M. V. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3756-3760 (1985).
- Schwartz, D. C. & Cantor, C. R. *Cell* **37**, 67-75 (1984).
- Carle, G. F. & Olson, M. V. *Nucleic Acids Res.* **12**, 5647-5664 (1984).
- Keesey, J. K., Bigella, R. & Fink, G. R. *J. Biol. Chem.* **254**, 7427-7433 (1979).
- Maher, L. J., Dervan, P. B. & Wold, B. *Biochemistry* **29**, 8820-8826 (1990).

ACKNOWLEDGEMENTS. We thank J. Hanish for helpful discussions. The work was supported by the National Institutes of Health and the Howard Hughes Medical Institute (predoctoral fellowship to S.A.S.).

Triple Helix Mediated Single-Site Enzymatic Cleavage of Megabase Genomic DNA

Scott A. Strobel and Peter B. Dervan*

Received _____

Division of Chemistry and Chemical Engineering
California Institute of Technology
Pasadena, CA 91125

*To whom correspondence should be addressed.

Oligonucleotide directed triple helix formation is a generalizable chemical approach for the recognition and cleavage of a single target site within several megabase pairs of duplex genomic DNA (1-4). Pyrimidine oligodeoxyribonucleotides 15 to 25 bases in length form a highly specific triple helix structure with purine tracts in double stranded DNA of high complexity (1-4). The pyrimidine oligonucleotide binds by Hoogsteen hydrogen bonding in the major groove of the DNA duplex parallel to the Watson-Crick purine strand (1). The recognition motif is generalizable to homopurine target sites utilizing thymine binding to adenine-thymine base pairs (T•A-T triplet) (5) and N3 protonated cytosine binding to guanine-cytosine base pairs (C+G-C triplet) (6-8). 5-bromouracil (BrU) and 5-methylcytosine (MeC) can be substituted in the third strand for T and C respectively, to generate oligonucleotides with higher binding affinities at slightly basic pH's (9). The triple helix motif is partially extended to mixed purine-pyrimidine sequences utilizing guanine binding to thymine-adenine base pairs (G•T-A triplet) (10) and alternate strand triple helix formation (11). More recently, a G-rich oligodeoxyribonucleotide third strand has been shown to bind antiparallel to the Watson-Crick purine strand via G•G-C, A•A-T, and T•A-T base triplets (12). The binding site size of triple helix recognition is sufficiently large to statistically identify a unique site in the human genome (>16 bp) (13, 14).

Koob et. al. demonstrated the feasibility of using a DNA binding protein to uniquely block the action of a methylase at a single overlapping recognition site while methylating all other sites (15, 16). Using this Achilles' heel cleavage

procedure, they demonstrated single site cleavage of yeast and *E. coli* genomes upon digestion with a restriction enzyme able to cut only at unmethylated sites (16). Protein-mediated Achilles heel cleavage may not, however, be readily generalizable to unique cleavage at selected genetic markers without artificial insertion of the target sequence due to the paucity of applicable DNA-binding proteins. A more general approach for recognition of endogenous DNA sequences might be a chemical approach based on oligonucleotide directed triple helix formation. This strategy offers a generalizable DNA binding motif that is capable of locally protecting the DNA from methylation at an overlapping target site while not affecting the activity of the methylase at its other recognition sites (17, 18). Triplex formation, global methylation and triple helix disruption results in DNA that is resistant to endonuclease digestion at all sites except the one previously bound by the oligonucleotide (17, 18) (Fig. 1). Subsequent restriction enzyme digestion produces nearly quantitative single site cleavage of large genomic DNA at endogenous sequences (4). In this chapter we fully describe the procedure for triple helix mediated enzymatic cleavage of the *Saccharomyces cerevisiae* genome at a single predetermined site on chromosome III (340 kb) using a 24 base oligonucleotide and EcoRI methylation and restriction enzymes.

Materials and Methods

Media, Solutions, and Reagents

YPD media: 1% (w/v) bactoyeast extract, 2% (w/v) bactopectone, 2% (w/v) glucose.

Autoclave yeast extract and peptone for 20 minutes, cool and add 1/25th volume of sterile 50% (w/v) glucose.

Spheroplasting Solution: 0.9 M sorbitol, 20 mM Tris pH 7.5, 50 mM EDTA pH 8.0, 7.5% β -mercaptoethanol, and 20 μ g/ml zymolyase 100T (ICN). Zymolyase is prepared as a stock solution at 2 mg/ml in 10 mM phosphate buffer pH 7.0 and 50% (w/v) glycerol.

Sucrose Solution: Same as above without zymolyase.

NDS Solution: 0.5 M EDTA pH 8.0, 10 mM Tris pH 7.5, 1% (w/v) lauryl sarcosyl, 1.0 or 0.5 mg/ml proteinase K (BRL).

Triple Helix/EcoRI Methylase Solution (THEM): 100 mM NaCl, 100 mM Tris-HCl, 10 mM EDTA, 2 mM spermine \cdot 4HCl (Nuclease free grade from Sigma).

Adjust final solution to pH 6.6, 7.0, 7.4, or 7.8.

Inactivation Buffer: 1% lauryl sarcosyl, 100 mM Tris-HCl, 10 mM EDTA. Adjust final solution to pH 9.0.

Wash Buffer: 10 mM Tris-HCl, 10 mM EDTA at final pH 9.0.

2xKGB Buffer: 200 mM potassium glutamate, 50 mM Tris-acetate, 20 mM magnesium acetate. Adjust to pH 7.6 with acetic acid or potassium hydroxide. Add BSA and β -mercaptoethanol to final concentration of 100 μ g/ml and 1 mM, respectively.

Preparation of Yeast Chromosomal DNA

1. Add 50 μ l of saturated overnight cultures of haploid yeast strain SEY6210 or recombinant derivatives (3, 4) to 100 mls of YPD media. Grow overnight to OD₆₀₀ between 4-6.
2. Harvest the cells by centrifugation at 5,000 rpm for 5 min. at 5° C. Discard the supernatant and wash with 20 mls 50 mM EDTA pH 8.0. Harvest and drain the pellet well.
3. Resuspend the yeast in 6-8 mls of spheroplasting solution and incubate at 37° C for 60 min. Gently swirl the solution every 15 min. to resuspend the cells. After spheroplasting is complete, centrifuge the cells as before, decant, and resuspend in 6-8 mls of sucrose solution. Add an equal volume of liquified 1.4% LMP Incert (FMC) agarose in sucrose solution preheated to 42° C, mix and cast in 6x6x100 mm molds. Solidify the agarose on ice to prevent leakage.
4. Carefully remove the plugs from the molds and transfer two plugs (=3.5 mls/plug) to 25 mls of NDS solution in a 50 ml conical centrifuge tube. Incubate with gentle shaking at 50° C for 24 hours. The plugs which were originally pale white become transparent upon lysis. Decant and add a second 25 ml aliquot of NDS solution with 0.5 mg/ml proteinase K and incubate at 50° C for an additional 3 hours.
5. Wash the plugs 3x30 mls for 20-30 min. with gentle shaking using 0.5 M EDTA 8.0. Repeat the 3x30 ml washes with 10 mM EDTA. Add 1/100 volume of 100 mM PMSF dissolved in 100% ethanol and incubate on shaker for at least 1 hour

to inactivate any residual proteinase K. Complete five 20-30 min. washes with 30 mls of 10 mM EDTA and store the plugs at 4° C until use.

Comments: DNA prepared from SEY6210 was more susceptible to degradation during preparation than other yeast strains tested. For this reason the standard procedure for DNA preparation from *S. cerevisiae* (19) was modified to maximize chromosomal integrity. In addition to the preparation in agarose blocks described above, DNA was prepared in agarose beads (16). Little difference was found in either the quality of the DNA or its reactivity in subsequent enzymatic treatments.

Preparation of Oligonucleotides

1. Synthesize oligonucleotides on a 1 μ mole scale using commercially available N,N diisopropyl β -cyanoethyl phosphoramidites (BrU and MeC phosphoramidites are available through Cruachem). Deprotect oligonucleotides in concentrated ammonium hydroxide at 55° C overnight and rotovap to dryness.
2. Purify oligonucleotides by 15% polyacrylamide gel electrophoresis in 1xTBE. Briefly visualize bands under a hand-held short-UV lamp and cut oligonucleotides from gel. Care should be taken with oligonucleotides containing BrU due to UV induced strand scission. Crush the polyacrylamide slice to powder and extract the oligonucleotide in 8.0 mls of 200 mM NaCl, 1 mM EDTA pH 8.0 at 37° C overnight. Pass extracted oligonucleotides through a 0.45 μ m cellulose acetate centrex filter (Schleicher and Schuell) to remove polyacrylamide fragments.

3. Desalt the oligonucleotides by slowly loading eluent on disposable C₁₈ reverse phase column (Waters). Remove salt by washing extensively with 20 mls of distilled water, and eluting in 2.0 mls of 40% acetonitrile in water. Determine concentration by OD₂₆₀ absorbance. Transfer oligonucleotides to eppendorf tubes in 20 nmole aliquots and dry on speedvac. Store at -20° C in the dark until use.

Triple Helix Mediated Enzymatic Cleavage

1. Cut 1.5 mm slices of yeast plugs ($\approx 50 \mu\text{l}$ volume) and transfer to 1.8 ml flat bottom microcentrifuge tubes. Wash 2 x 1.0 ml with THEM buffer for 15 minutes. Decant and overlay the plug with 120 μl of THEM buffer.
2. Add 2 μl of 100 μM oligonucleotide to the overlay and incubate at room temperature on shaker (150 rpm) for 8 hours to facilitate oligonucleotide diffusion and triple helix formation.
3. Add 1 μl of 20 mg/ml acylated BSA (nuclease free grade from Sigma), 1 μl of 32 mM S-adenosyl-methionine in 5 mM H₂SO₄, 10% ethanol, and 2 μl of 40 units/ μl EcoRI methylase (New England Biolabs). Incubate an additional 3-5 hours at room temperature with shaking.
4. Simultaneously disrupt the triplex and inactivate the methylase by removing the supernatant above the plug and adding 1.0 mls of inactivation buffer. Heat plugs to 55° C for 30 min. Place tubes on ice following inactivation to harden the agarose plug before decanting.

5. Remove lauryl sarcosyl, oligonucleotide and methylase from plug by washing 4 x 15 min. with 1.0 ml of wash buffer.
6. Equilibrate the DNA in restriction enzyme buffer by washing 2 x 15 min. with 1.0 ml of 2xKGB. Decant and overlay with 120 μ l of buffer. Add 1 μ l of 20 units/ μ l EcoRI restriction enzyme (New England Biolabs) and digest DNA for 6 hours at room temperature on shaker.
7. Decant plugs and add 1.0 ml of 0.5x TBE. Disrupt any DNA protein interactions by heating to 55° C for 10 min. and cooling quickly on ice.
8. Resolve cleavage products by pulsed field gel electrophoresis (20, 21) using a 0.5x TBE, 1% agarose gel at 14° C, 200 V on a CHEF II system (BioRad). Ramp switch times from 10-40 sec. for 18 hrs. and 60-90 sec. for 6 hrs. Visualize products by ethidium bromide staining (500 ng/ml) and Southern blotting with 0.45 μ m Nytran membranes according to manufactures protocols (Schleicher and Schuell).

Comments: The incubation time necessary for complete methylation is dependent upon the pH of the buffer. The pH optimum is 8.0 for EcoRI methylase, but triple helix formation is more effective at neutral pH. Methylation is complete within three hours at pH 7.4, but 4.5 hours are necessary at pH 6.6. Thus, the methylation time must be optimized for the conditions chosen.

Commercial preparations of EcoRI methylase contain endonuclease impurities that are active in the presence of magnesium. Fortunately, methylases do not require a metal cofactor for activity. To prevent nonspecific degradation of the chromosomal DNA, the methylation reactions were performed without

magnesium in the presence of the metal chelator EDTA. After methylation is complete, the methylase impurities must be inactivated in warm detergent before addition of the restriction enzyme buffer.

Because triple helix formation can specifically block the activity of both methylases and restriction enzymes (17, 18) it is necessary to remove the oligonucleotide from the target site. Utilizing the observation that triple helix formation is temperature, polycation and pH dependent (1, 3) the complex can be readily disrupted in the absence of spermine at moderate temperatures and high pH. After triplex disruption and before the addition of magnesium (a dication necessary for the restriction enzyme that also stabilizes triple helix formation) the plug must be thoroughly washed to remove the oligonucleotide.

Two Dimensional Pulsed Field Gel Electrophoresis

1. Follow steps 1-3 as described above. After complete methylation of the DNA but before restriction enzyme digestion, separate the yeast chromosomes by pulsed field gel electrophoresis in a 0.5x TBE, 0.8% Incert LMP agarose gel at 14° C and 200 V. Set switch times at 60 sec. for 18 hours and 90 sec. for 6 hours to fully resolve all the yeast chromosomes. Include an untreated control lane to serve as a size standard.
2. Cut the size standard from the gel and stain with ethidium bromide. Do not stain the methylated DNA. Based upon the location of the chromosomal bands in the standard lane, carefully cut a full length strip of methylated DNA and

place it in a 15 ml centrifuge tube containing 4 mls of 2x KGB buffer. Equilibrate for 10 min. at room temperature.

3. Decant and add 4 mls of 2x KGB buffer and 10 μ l of EcoRI restriction enzyme (200 units). Digest at RT with shaking (60 rpm) for 6 hrs.
4. Load strip onto a second 1% agarose pulsed field gel containing 0.5x TBE at 14° C and 200 V. Set switch times as in step 1. Visualize products by ethidium bromide staining.

Results and Discussion

Studies with plasmid DNA demonstrate that oligonucleotide directed triple helix formation in combination with appropriate methylases and restriction enzymes can limit the operational specificity of restriction enzymes to endonuclease recognition sequences that overlap oligonucleotide binding sites (17, 18). To demonstrate the feasibility of single site enzymatic cleavage of genomic DNA by triple helix mediated Achilles' heel cleavage, a sequence containing an overlapping 24 bp purine tract and 6 bp EcoRI site was inserted proximal to the *LEU2* gene (22) on the short arm of chromosome III (340 kb) by homologous recombination (Fig. 2) (3). Oligonucleotides designed to form triple helix complexes that overlap half of the EcoRI recognition site were synthesized with CT, MeCT, MeCBrU nucleotides (Fig. 2) (9). The genetic map of yeast chromosome III (23, 24) and affinity cleaving data (3) indicate that cleavage at the target site should produce two fragments 110 ± 10 and 230 ± 10 kb in size (Fig. 2).

Use of the MeCT oligonucleotide for triple helix formation at pH 7.4 followed by EcoRI methylase and endonuclease treatments, resulted in exclusive cleavage of chromosome III at the target site (Fig. 3, lane 1). No cleavage was detected on any other chromosomes under these conditions. Cleavage was neither observed in the absence of oligonucleotide (Lane 2) nor in a yeast strain lacking the target sequence (Lane 3). The expected 110 kb product was visualized with ethidium bromide staining and confirmed by Southern blotting with a *HIS4* marker (Fig. 3C) (25). The 230 kb product comigrated with chromosome I, but was detected by Southern blotting with a *LEU2* marker (Fig. 3B) (22). The cleavage efficiency was $94 \pm 2\%$ with a three hour methylase incubation. Longer methylation times resulted in a gradual reduction in the observed cleavage efficiency, suggesting that the oligonucleotide dissociation rate might be the limiting factor for efficiency in this system (26).

The specificity of triple helix formation has been shown to be pH dependent (1, 2, 3). At lower pH, sequences of near but imperfect similarity can be bound and subsequently cleaved. In agreement with this observation, secondary cleavage sites were revealed as a function of oligonucleotide composition and pH (Fig. 4). Using either the MeCT or MeCBrU oligonucleotides at or below pH 7.4, 170 and 650 kb secondary cleavage products were observed (Fig. 4, lanes 5 and 9). The cleavage site can be assigned to chromosome II (820 kb) by two-dimensional pulsed-field gel electrophoresis (Fig. 5). Additional secondary cleavage sites were observed with the MeCBrU oligonucleotide at pH 6.6 (Fig. 4, lane 9). The 180 and 480 kb products were assigned to chromosome XI (660 kb), and the 330 and 780 kb products were assigned

to the VII/XV doublet (1100 kb) (Fig. 5). Cleavage at all secondary sites could be eliminated at a threshold pH for each oligonucleotide (Fig. 4, lanes 2, 8, and 12). Because of the greater triple helix stability at pH's optimal for the methylase (pH 8.0), MeC substituted oligonucleotides are preferred over C substitutions in this assay. BrU substitutions offered little additional pH stability and increased the number of secondary cleavage sites observed. This property of BrU substituted oligonucleotides is useful in searching for endogenous homopurine-methylase sites in unsequenced DNA by using degenerate oligonucleotides containing MeC and BrU.

Many methylases, including AluI and dam, exhibit optimal activity in low salt buffers. Unfortunately, spermine rapidly precipitates DNA at low sodium chloride concentrations. To prevent precipitation during triple helix formation and methylation reactions, the salt concentration must be adjusted to 25-50 mM. The enzyme activity retained under these conditions is sufficient to yield complete methylation of genomic DNA if adequate time and enzyme concentration are employed.

Not all methylases are compatible with this technique. BamHI methylase is significantly inhibited by spermine. HaeIII restriction enzyme can cut hemimethylated sites, making it much more difficult to achieve complete methylation (27). An additional concern when working with genomic DNA, particularly with enzymes sensitive to CpG methylation such as HpaII and HhaI, is the prior methylation state of the target DNA. A large yet variable percentage of CpG sequences in mammalian cells are methylated *in vivo* (28, 29). Achilles heel

cleavage at these sites would not be possible unless methylation free DNA could be prepared.

Affinity cleaving using oligonucleotide-EDTA•Fe, a technique that generates cleavage at all sites of oligonucleotide binding but in low yield, demonstrated that triplex formation is occurring at several positions on the yeast chromosomes besides the single target site (3). Triple helix mediated Achilles heel cleavage exposes only those sites that are tightly bound and partially overlap a complete methylase/endonuclease site (4, 17, 18). The probability of finding an endogenous homopurine sequence that overlaps a EcoRI methylase site is quite low, possibly one every 500 kb or less (3). The frequency of targetable sequences is increased, however, when other methylase/endonuclease pairs are considered. In addition to TaqI and EcoRI (4, 17, 18), single site protection of plasmid DNA has been achieved with MspI, HpaII, and AluI methylase/endonucleases pairs. Several other methylases could be used, but remain untested. To date, complete methylation of yeast genomic DNA under conditions compatible with triple helix formation has been achieved with EcoRI, AluI and dam methylases. The list will expand as target sites of interest are identified. Although all enzyme sets tested might not be effective, the use of several different methylase/endonuclease combinations could increase the frequency of cleavable sites in endogenous sequences to as high as one every 10-30 kb (3, 27).

The generalizability of triple-helix mediated enzymatic cleavage affords high specificity that can be readily customized to unique genetic markers without artificial insertion of a target sequence. Extensive sequencing to identify target sites

could be avoided by using degenerate oligonucleotides to screen genetic markers for overlapping triple helix/endonuclease sites. The potential generalizability of triple helix mediated Achilles heel cleavage, a technique capable of near quantitative cleavage at a single site in at least 14 megabase pairs of DNA, could assist in physical mapping of chromosomal DNA and expedite isolation of DNA segments linked to disease.

Acknowledgement. We are grateful to Dr. John Hanish for helpful discussions, the National Institute of Health for grant support, and the Howard Hughes Medical Institute for a predoctoral fellowship to SAS.

Fig. 1 General scheme for single-site enzymatic cleavage of genomic DNA by oligonucleotide-directed triple helix formation (4). Chromosomal DNA is equilibrated with an oligodeoxyribonucleotide in a methylase compatible buffer containing polycation. EcoRI methylase, which methylates the central adenines of the sequence 5'-GAATTC-3' and renders the sequence resistant to cleavage by EcoRI restriction endonuclease, is added and allowed to methylate 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 reequilibrated in restriction enzyme buffer and cut to completion with EcoRI restriction endonuclease. The cleavage products are separated by pulsed-field gel electrophoresis and efficiencies quantitated by Southern blotting (From ref. 4).

Fig. 2 (Left) Genetic map of *S. cerevisiae* chromosome III. The location of the *HIS4* and *LEU2* loci (boxes), the centromere (circle), and the triple helix-EcoRI target site are indicated. The expected sizes of the cleavage products are shown. **(Right)** Schematic diagram of the triple helix complex overlapping the EcoRI methylase-endonuclease site. The pyrimidine oligonucleotide is bound in the major groove parallel to the purine strand of the DNA duplex, and covers half of the EcoRI site (From ref. 4).

Fig. 3 Single site enzymatic cleavage of yeast genomic DNA with reagents indicated above figure. Products were visualized by ethidium bromide staining (A) and Southern blotting with *LEU2* (B) and *HIS4* (C) chromosome markers (From ref. 4).

Fig. 4 Triple helix mediated enzymatic cleavage of the yeast genome as a function of

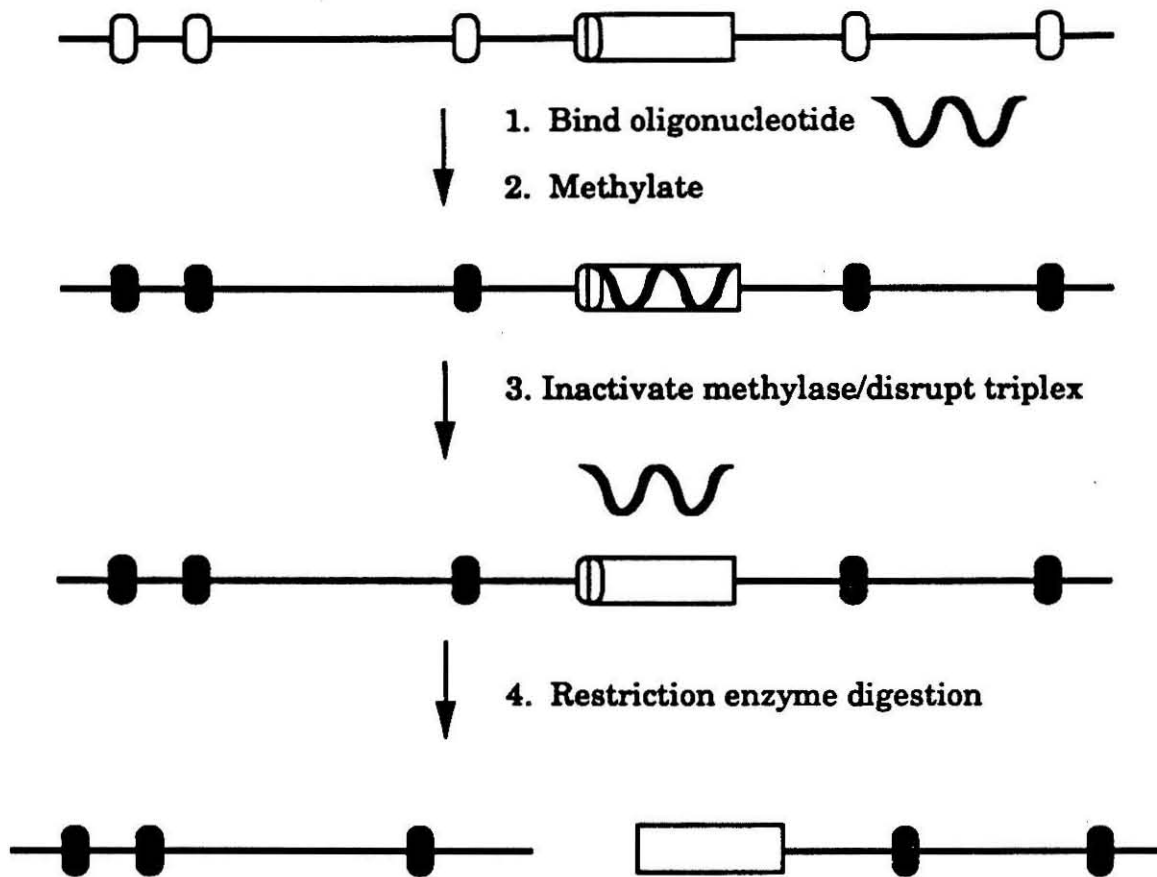
oligonucleotide composition and pH. (A) Lane 1-12: Reaction on a yeast strain containing the triple helix target site with oligonucleotides (CT, MeCT, and MeCBrU) and pH (6.6, 7.0, 7.4, and 7.8) as indicated above figure (From ref. 4).

Fig. 5 Two dimensional pulsed field gel analysis of MeCBrU oligonucleotide incubated at pH 6.6. Intact chromosomes form a diagonal with cleavage products located immediately to the left of the chromosome of origin.

1. H. E. Moser, P. B. Dervan, *238*, 645-650 (1987).
2. S. A. Strobel, H. E. Moser, P. B. Dervan, *J. Am. Chem. Soc.* **110**, 7927-7929 (1988).
3. S. A. Strobel, P. B. Dervan, *Science* **249**, 73-75 (1990).
4. S. A. Strobel, P. B. Dervan, *Nature* **350**, 172-174 (1991).
5. G. Felsenfeld, D. R. Davies, A. Rich, *J. Am. Chem. Soc.* **79**, 2023-2024 (1957).
6. M. N. Lipsett, *J. Biol. Chem.* **239**, 1256-1260 (1964).
7. P. Rajagopal, J. Feigon, *Biochem.* **28**, 7859-7870 (1989).
8. C. de los Santos, M. Rosen, D. Patel, *Biochem.* **28**, 7282-7289 (1989).
9. T. J. Povsic, P. B. Dervan, *J. Am. Chem. Soc.* **111**, 3059-3061 (1989).
10. L. C. Griffin, P. B. Dervan, *Science* **245**, 967-971 (1989).
11. D. A. Horne, P. B. Dervan, *J. Am. Chem. Soc.* **112**, 2435-2437 (1990).
12. P. A. Beal, P. B. Dervan, *Science* **251**, 1360 (1991).
13. P. B. Dervan, in *Nucleic Acids and Molecular Biology* F. Eckstein, D. M. J. Lilley, Eds. (Springer-Verlag, Heidelberg, 1988), vol. 2, pp. 49-64.
14. P. B. Dervan, in *Oligonucleotides as Inhibitors of Gene Expression* J. S. Cohen, Eds. (MacMillan Press Ltd., London, 1990) pp. 197-210.


15. M. Koob, E. Grimes, W. Szybalski, *Science* **241**, 1084-1086 (1988).
16. M. Koob, W. Szybalski, *Science* **250**, 271-273 (1990).
17. L. J. Maher, B. Wold, P. B. Dervan, *Science* **245**, 725-730 (1989).
18. J. C. Hanvey, M. Shimizu, D. Wells, *Nuc. Acids Res.* **18**, 157-161 (1990).
19. C. L. Smith, S. R. Klco, C. R. Cantor, in *Genome Analysis: A Practical Approach* K. Davies, Eds. (IRL Press, England, 1988) pp. 41-72.
20. D. C. Schwartz, C. R. Cantor, *Cell* **37**, 67-75 (1984).
21. G. F. Carle, M. V. Olson, *Nuc. Acids Res.* **12**, 5647-5664 (1984).
22. A. Andreadis, Y. P. Hsu, G. B. Kohlhow, P. Shimmel, *Cell* **31**, 319-325 (1982).
23. R. K. Mortimer, D. Schild, *Microbiol. Rev.* **49**, 181-212 (1985).
24. G. F. Carle, M. V. Olson, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3756-3760 (1985).
25. J. K. Keeseey, R. Bigelis, G. R. Fink, *J. Biol. Chem.* **254**, 7427-7433 (1979).
26. L. J. Maher, P. B. Dervan, B. Wold, *Biochemistry* **29**, 8820-8826 (1990).
27. C. Kessler, H. J. Holtke, *Gene* **47**, 1-153 (1986).
28. A. Bird, *Nature* **321**, 209-213 (1986).


29. A. Bird, M. Taggart, M. Frommer, O. J. Miller, D. Macleod, *Cell* **40**, 91-99 (1985).

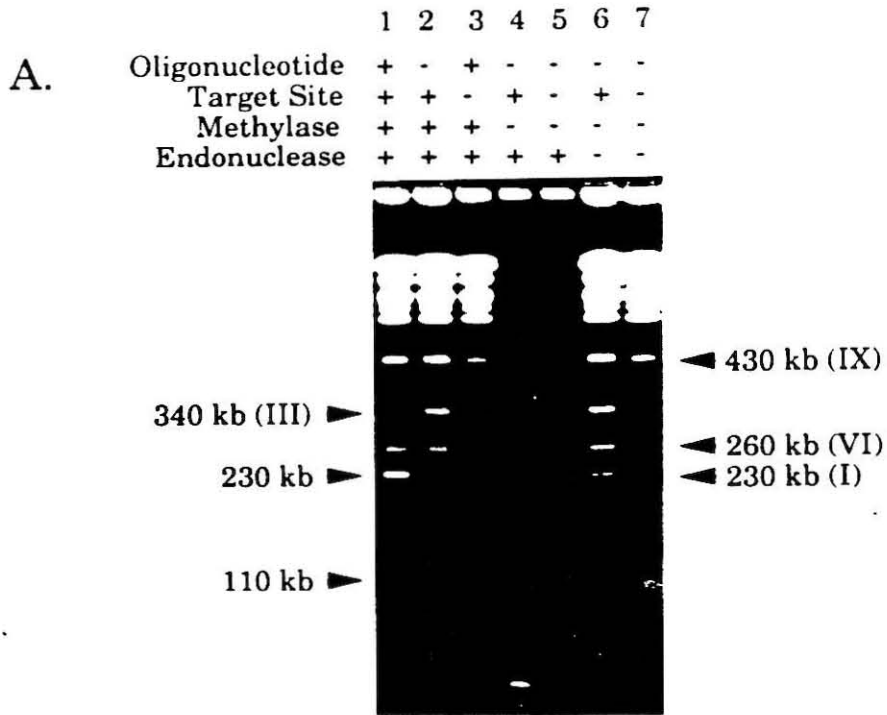


 - Restriction enzyme site, cleavable

 - Methylated restriction enzyme site, not cleavable

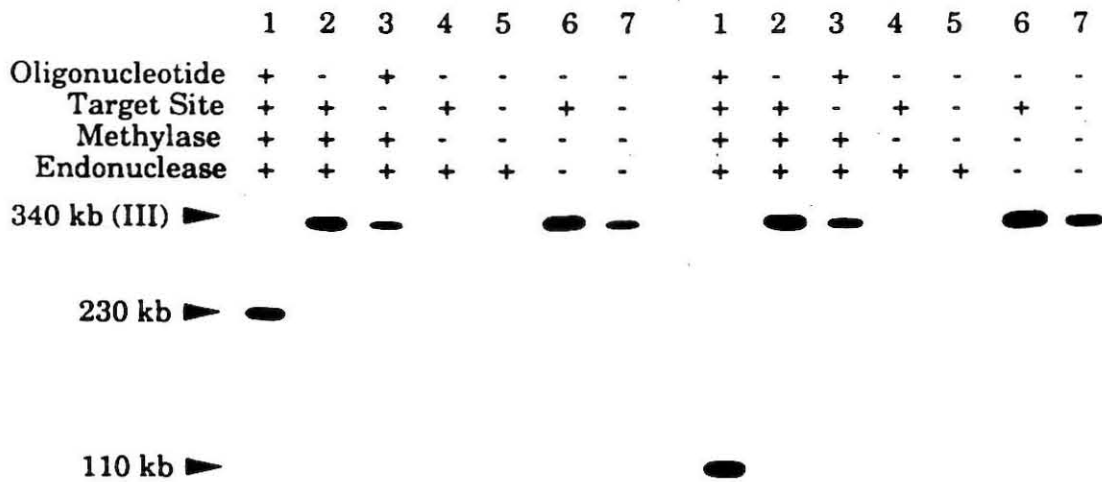
 - Oligonucleotide binding site

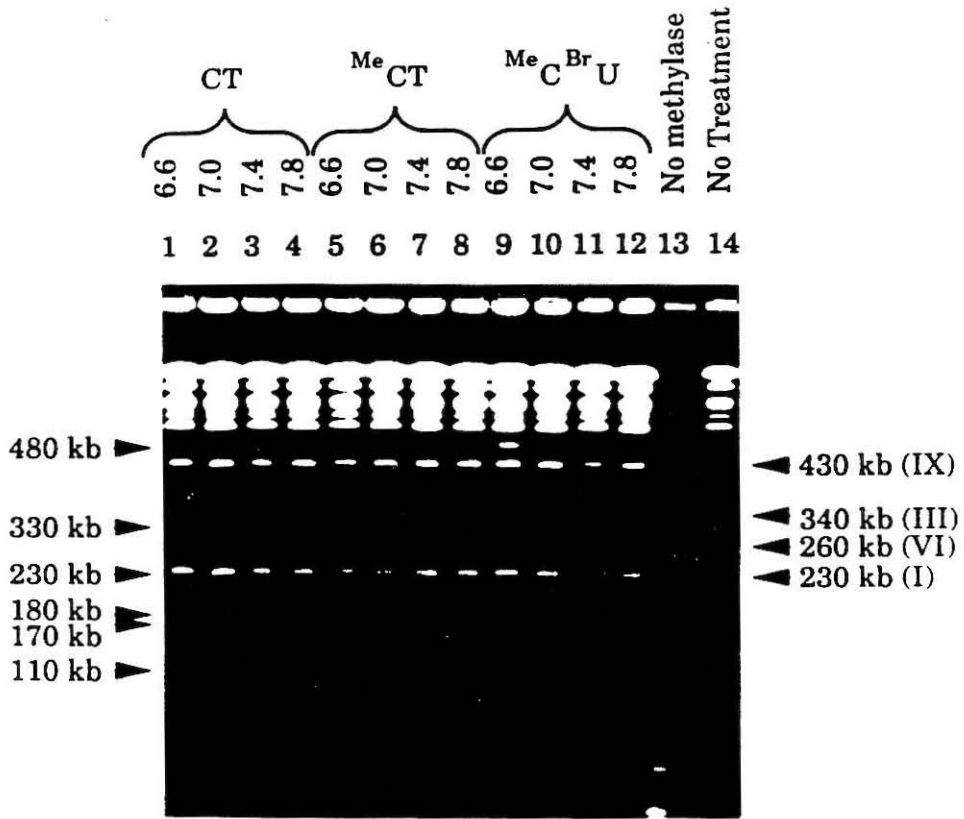
 - Oligonucleotide

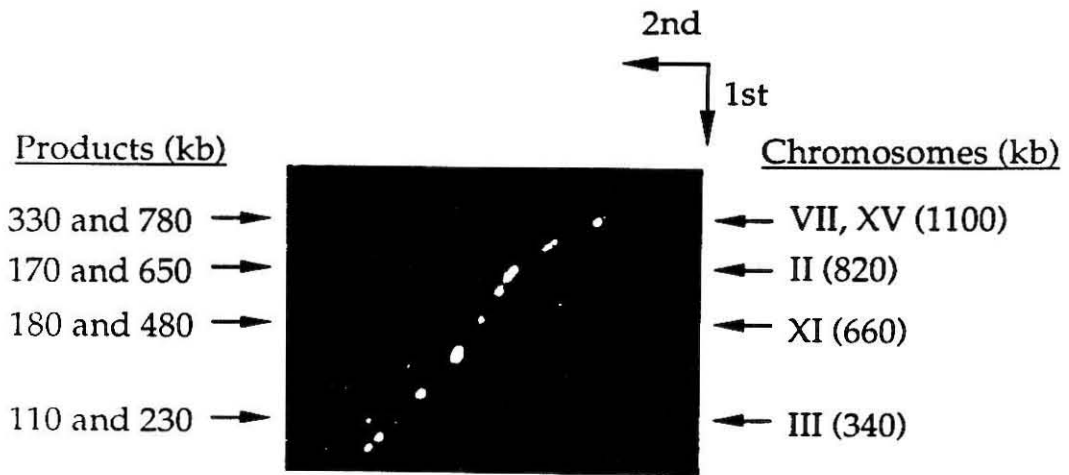


B. *LEU2*

C. *HIS4*







Conclusion. Single site cleavage of the yeast genome was achieved enzymatically by oligonucleotide directed triple helix formation. Triple helix specificity alone was insufficient to define one site on a single yeast chromosome even under optimized conditions. By coupling the specificities of triple helix formation and enzymatic methylation/digestion, a single cleavage site was detected within plasmid and yeast genomic DNA. The sequence requirement of a triplex binding site overlapping a complete methylation site was sufficient to define a single site within 14 Mb of DNA and cut it at efficiencies exceeding 94%. Secondary target sites within the yeast genome could be detected with this cleavage technique through the use of modified oligonucleotides at low pH. The number of cleaved sites were surprisingly few considering the extent of degradation observed with MeCBrU oligonucleotides by affinity cleaving. The chromosomal identification and sizes of the products generated by cleavage at the less stringent triple helix condition were determined by two-dimensional pulsed-field gel electrophoresis.

Materials and Methods.

Materials Restriction enzymes were obtained from New England Biolabs, Boehringer Mannheim, and Gibco BRL (SstI only). All methylases were obtained from New England Biolabs and used without further purification. T4 DNA ligase and the Klenow fragment of *E. coli* DNA

polymerase were obtained from New England Biolabs. Sequenase 2.0 dideoxy sequencing kits were purchased from Unites States Biolabs. Taq DNA polymerase and PCR buffers were obtained from Perkin Elmer Cetus. PCR reactions were performed on a Perkin Elmer Cetus thermocycler. Prep-a-Gene DNA purification matrix and buffers were obtained from Bio-Rad. Large scale plasmid purification was performed using maxi-plasmid purification kits from Qiagen according to manufacture suggested protocols. Radioactive nucleotides $^{32}\text{P-}\alpha\text{-dCTP}$ and $^{32}\text{P-}\alpha\text{-dATP}$ were obtained from Amersham [10 mCi/ml, 3000 Ci/mmol]. Unlabeled mononucleotides were purchased from Pharmacia, mixed at 2.5 mM in each nucleotide, and stored in small aliquots for PCR reactions. Sephadex G-10 and G-50 matrices, spermine•4HCl (nuclease free), ampicillin, lauryl sarcosyl, isopropyl $\beta\text{-D}$ -thiogalactopyranoside (IPTG), 10 mg/ml salmon sperm DNA (fragmented and phenol extracted), and bovine serum albumin (20 mg/ml nuclease free) were obtained from Sigma. 5-Bromo-4-chloro-3-indolyl $\beta\text{-D}$ -galactopyranoside (X-gal) and proteinase K were purchased from Boehringer Mannheim. Zymolyase was obtained from ICN Biochemicals. $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ and ultrapure dithiothreitol were obtained from Baker Chemical and Gibco BRL, respectively. YNN295 yeast chromosomal size standards were a gift from Dr. Bruce Birren. The concatenated λ DNA size standard and Incert low melting point (LMP) agarose was purchased from FMC Bioproducts. Electrophoresis grade agarose and LMP agarose were purchased from Gibco BRL.

Plasmid pUC19 was obtained from Pharmacia. Plasmid YEp13 (28) and the haploid yeast strain SEY6210 (30) were gifts from the laboratory of Professor Scott Emr. Plasmid YCp503 was a gift from Professor Randy Scheckman (24). Plasmid pMA91 was the gift of Professor Judy Campbell (25).

Chromosomes and cleavage products were resolved on a CHEF-DRII Pulsed-field gel electrophoresis chamber from Bio-Rad. DNA blotting was accomplished using a Stratagene Pressure Blotter. 0.45 Micron Nytran membrane was obtained from Schleicher and Schuell. DNA nicking was accomplished with a UV Products short UV TS-20 transilluminator. The DNA was immobilized by UV irradiation with a Stratagene Stratalinker. Radioactive emissions were detected on a Beckman LS 3801 liquid scintillation counter. DNA hybridization (Southern blotting) was accomplished in a Robins Scientific rotating hybridization incubator, Model 310. Autoradiography was performed using Kodak X-Omat AR film with an intensifying screen at -70° C and developed on a Kodak M35 A X-OMAT film processor. Densitometric traces were recorded on an LKB Bromma Ultrascan XL Laser Densitometer.

Media and Buffers. YPD media was prepared by mixing 10 g yeast extract, 20 g bactopectone, 20 g bactoagar (if for plates), 960 mls of water and autoclaving for 20 min. 40 mls of sterile 50% w/v glucose was added to bring the volume to 1.0 L.

Minimal media plates (leu⁻) were prepared by mixing 6.7 g yeast nitrogen base, and 25 g bactoagar in 860 mls of water and autoclaving for 20 min. Sterile 50% w/v glucose (50 mls) and sterile 10x amino acid mix (200 mg uracil, 200 mg histidine, 200 mg adenine, 300 mg lysine, 200 mg tryptophan, 200 mg methionine, 300 mg tyrosine in 1000 ml sterilized for 20 min. in autoclave) (100 mls) were added, mixed, and poured into plates.

LB media and TYE plates were prepared as previously described (39). Stock solutions for DNA electrophoresis and hybridization were as follows:

-1xTAE, 40 mM tris-acetate, 1 mM EDTA, pH adjusted to 8.0 with glacial acetic acid.

-1xTBE, 45 mM tris-base, 45 mM boric acid, 2 mM EDTA, pH 8.0.

-20xSSPE, 3.6 M NaCl, 0.2 M NaPO₄, 20 mM EDTA, pH 7.7.

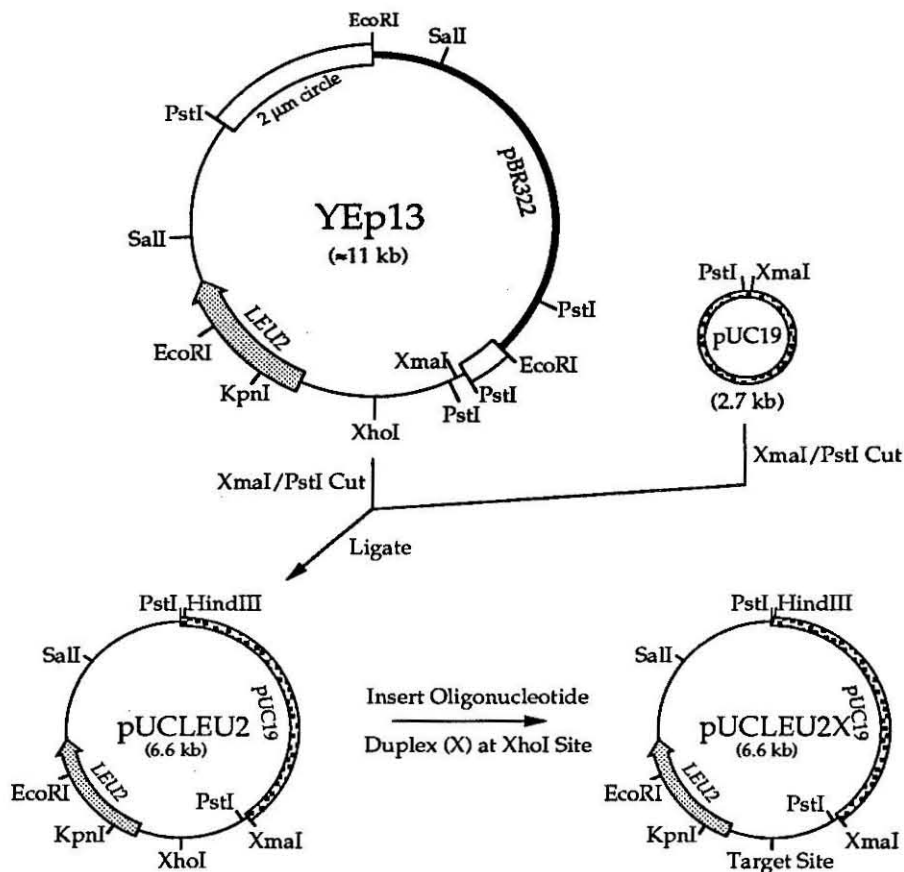
-20xSSC, 3.0 M NaCl, 0.3 M NaCitrate pH 7.0.

-50xDenhardt's Reagent, 1% w/v ficoll, 1% w/v polyvinylpyrrolidone, 1% bovine serum albumin.

Oligonucleotide Synthesis and Purification. Oligonucleotides for triple helix formation and affinity cleaving were synthesized, deprotected, and purified as described in Chapter 2. Oligonucleotides for use in plasmid construction, PCR, or as sequencing primers were synthesized on an Applied Biosystems 380 B DNA synthesizer and deprotected in concentrated NH₄OH at 55° C for 12 to 24 hours in screw cap eppendorf tubes. Following deprotection the ammonia was removed with argon for 30-60 min. The oligonucleotides were dried by lyophilization, redissolved in 1.0 ml water, desalted through a Sephadex G-10 size exclusion spin column, and the oligonucleotide concentration determined by λ_{260} absorbance.

Plasmid Construction. Target site insertion into yeast chromosome III was achieved by homologous recombination with a *LEU2* fragment containing the target sequence. The *LEU2* fragment was derived from the plasmid YEp13, which contained the gene on a PstI fragment inserted into the shuttle vector CV4, a pBR322 derivative containing the small EcoRI fragment from the 2 μ circle (12, 28, 29). Although YEp13 contained a unique XhoI site for target site insertion, it also contained two undesirable features: (i) it was a large low copy number plasmid, difficult to prepare in good yield, and (ii) it contained a 2 μ circle replication origin that facilitated extrachromosomal expression of the *LEU2* phenotype. Because this could potentially reduce the recombination frequency of the target site into chromosome III, a high copy

Fig. 3.7. Construction of plasmid pUCLEU2 and its derivatives from parent plasmid YEp13. **Top.** The PstI/XmaI fragment of YEp13 was ligated into the polylinker region of pUC19 by a three piece ligation. The resulting plasmid, pUCLEU2, contained a unique XhoI site upstream of the LEU2 gene into which oligonucleotide duplexes A, B, C, D, E, H, and P were inserted. The plasmid also contained a unique HindIII site, opposite the XhoI site, that was frequently used for plasmid linearization in cleavage reactions. **Bottom.** The plasmid names, duplex sequences, and restriction enzyme sites inserted into the parent plasmid pUCLEU2. An additional plasmid containing two inserts of the A duplex was also constructed (pUCLEU2A₂). Constructs E, H, and P are from sequences in human chromosome 4 (see Chapter 4). The sequence in plasmid P was changed slightly from the endogenous sequence to generate a BamHI site. The endogenous sequence reads GGATCT instead of GGATCC. Both sites are recognized by the dam methylase sensitive enzyme MflI.



PLASMID	DUPLEX INSERT (X)	ENZYMES	
		5' End	3' End
pUCLEU2A	5'-TCGACGAATTCTTTTCTTTCTTTCTTTCTTTCTTCGAG-3' 3'-GCTTAAGAAAAAGAAAGAAGAAAAGAAAGCTCAGCT-5'	EcoRI	TaqI
pUCLEU2B	5'-TCGACGGATCCTTTTCTTTCTTTCTTTCTTTCCGGG-3' 3'-GCCTAGAAAAAGAAAGAAGAAAAGAAAGGCCAGCT-5'	dam/MboI BamHI	MspI HpaII
pUCLEU2C	5'-TCGACTTTTCTTTCTTTCTTTTCTTTACTAGTAAAAGAAAAGAAAAGAAAAG-3' 3'-GAAAAGAAAAGAAAAGAAAATGATCATTTTCTTTCTTTCTTTTCAGCT-5'	Alkylation	
pUCLEU2D	5'-TCGACTAGTAAAAGAAAAGCTTTCTTTCTTTACTAG-3' 3'-GATCATTTCTTTTTCGAAAAGAAGAATGATCAGCT-5'	Alkylation	
pUCLEU2E	5'-TCGACTCGCTAATGGAAAGAGAGAGAGCTCACGTG-3' 3'-GAGCGATTACCTTTCTCTCTCTCGAGTGCACAGCT-5'	-	AluI SstI
pUCLEU2H	5'-TCGACACAGAAGGATATAAAGAAAGCTTTAG-3' 3'-GTGTCTTCCTATATTTCTTTTCGAAATCAGCT-5'	-	AluI HindIII
pUCLEU2P	5'-TCGACATAGGGGAGAGGCAGGGGATCCGTG-3' 3'-GTATCCCCTCTCCGTCCCTAGGCACAGCT-5'	-	dam/MboI MfiI BamHI*

number plasmid, pUCLEU2, lacking the 2 μ circle was constructed (Fig. 3.7).

Restriction maps of YEp13 showed the *LEU2* insert to be on a single PstI fragment with an XmaI site immediately adjacent to one end. YEp13 DNA was cut with PstI and XmaI and ligated into PstI/XmaI linearized pUC19 DNA (Fig. 3.7). DNA was transformed into XL-1 blue competent cells and transformants selected by α -complementation on TYE plates with ampicillin, IPTG, and X-gal. Few white colonies were detected if the YEp13 fragment was gel purified, and many plasmids with small XmaI/PstI inserts were detected if the DNA from YEp13 was not purified after digestion. This suggested that there might be a second PstI site internal to the XmaI site, and that the desired product required a three piece ligation. Plasmid DNA isolated from several clones was screened by restriction enzyme digestion (PstI, XhoI, EcoRI, Sall, XmaI, and HindIII) to identify a single colony with the proper construction. Sequencing of this plasmid revealed a PstI site approximately 60 bp from the XmaI site indicating that it was the product of a three piece ligation. Despite the difficulty of constructing the desired product because of an error in the restriction map, PstI digestion of this plasmid, pUCLEU2, generated a single fragment complementary to the defective *leu2* locus on yeast chromosome III of SEY6210.

pUCLEU2 contained a single XhoI site upstream of the *LEU2* gene into which oligonucleotide duplexes with homopurine target sites were ligated (Fig. 3.7). A number of duplexes with target sites for affinity cleaving (A and B), alkylation (C and D), and enzymatic cleavage (A, B, E, H, and P) were ligated into the XhoI site of pUCLEU2 (Fig. 3.7). The duplexes were designed to contain ends compatible with ligation into a XhoI site, and an internal C to destroy the XhoI site upon duplex insertion. In this way, the desired plasmid could be selected prior to transformation by performing a second XhoI digest

after ligation, relinearizing only those plasmids lacking an insert. The internal C also generated a Sall site if multiple duplexes were inserted, facilitating rapid clone identification by Sall digestion following transformation and colony selection. In this way, plasmid pUCLEU2A₂ was selected containing two tandem inserts of the A duplex.

Molecular cloning was performed using techniques previously described (39). pUCLEU2/duplex ligations were performed with 1 unit T4 DNA ligase using a 20-fold molar ratio of unphosphorylated duplex ($\approx 5 \times 10^{12}$ moles) to plasmid ($\approx 2.5 \times 10^{13}$ moles or 1 μ g) in 20 μ l. Ligations were performed in 1xKGB buffer (100 mM potassium glutamate, 25 mM tris-acetate, 10 mM magnesium-acetate, 50 μ g/ml BSA, 0.5 mM β -mercaptho-ethanol, pH 7.6) and 0.5 mM ATP at 16° C for 2-4 hours, after which XhoI (20 units) was added directly to the reaction mixture and incubated at 37° C for an additional hour. Transformation, selection, restriction enzyme digestion, and plasmid isolation were performed as described. Transformants were characterized by restriction enzyme digestion of purified plasmid DNA. The orientation and sequence of the inserts was directly determined by dideoxy sequencing with Sequenase 2.0. Large scale plasmid preparation was performed by double cesium chloride banding (pUCLEU2, A, A₂ and B) or ion exchange chromatography (pUCLEU2C, D, E, H, and P). For reasons that are not clear, control ligations lacking the duplex generated significantly more transformants than ligations with the duplex, despite digesting both with XhoI. In spite of this troubling observation, roughly half of the resulting colonies contained the desired insert.

Homologous Recombination of Yeast Chromosome III. Competent yeast were transformed with the PstI fragment from pUCLEU2 and its derivatives to homologously recombine the triple helix target site into

chromosome III (40-43). Haploid yeast strain SEY6210 (*MAT α leu2-3, 112 ura3-52 his3- Δ 200 trp- Δ 901 lys2-801 suc2- Δ 9 GAL*) was grown in 100 mls YPD at 30° C to $A_{600} = 0.6$ to 0.8 (OD readings taken on 10 fold dilutions from the concentrated culture). Cells were harvested by centrifugation at 5000 rpm for 5 min. at 5° C, decanted, washed in sterile water, reharvested, and decanted. The cells were then resuspended in 20 mls 0.1 M lithium-acetate, 5 mM tris-HCl pH 8.0, 0.1 mM EDTA and incubated at 30° C for 30 min. The cells were harvested, resuspended in 1 ml of lithium-acetate solution, and 50 μ l transferred to 1.8 ml eppendorf tubes for transformation.

Plasmid DNA (5 μ g) was digested to completion with PstI, added to the competent yeast cells, and the mixture incubated at 30° C for 30 min. Freshly sterilized 40% polyethylene glycol-4000 (0.6 ml) was added, mixed and incubated an additional hour at 30° C. Cells were heat shocked at 37° C for 5 min., collected by quickly spinning in a microcentrifuge, decanted, resuspended in 125 μ L water, and plated on leucine deficient minimal media. Transformants were incubated for 48 hours at 30° C, after which several very small colonies and 10-25 large colonies were visible. Replating the small colonies on minimal media showed they were defective for growth, while the large colonies showed significant growth after 2 additional days at 30° C. Four of the large colonies were picked for storage and one of them screened for proper insertion of the target site.

Confirmation of the Target Site in Yeast Chromosome III. Two procedures were used to demonstrate the target site had been inserted at the desired location within the yeast genome: (i) DNA hybridization with the *LEU2* gene (39, 44), and (ii) PCR amplification of the target site region followed by product digestion with informative restriction enzymes (45-48). If the target site was properly inserted, only one copy of the *LEU2* gene would be

present and hybridization would show its location on chromosome III. If recombination had occurred at a second nonhomologous site, the hybridization would detect two chromosomes containing the *LEU2* gene, the parental chromosome III and a second chromosome that had undergone recombination.

Chromosomal DNA was prepared in Incert LMP agarose from recombinant yeast strains as described (14, 16, 22). The chromosomes were resolved by pulsed field gel electrophoresis using a 1% agarose, 0.5xTBE gel, 6.0 V/cm, 120° switch angle, 60 sec. switch times for 16 hours, followed by 90 sec. switch times for 8 hours at 14° C (3-8). The DNA was stained with ethidium bromide, photographed, nicked on a short UV light box for 30 sec. each side, and denatured by incubating the gel in 300 mls of 0.4 M NaOH/0.6 M NaCl twice for 30 min. The DNA was transferred to a 0.45 µm Nytran membrane by pressure blotting (75 torr) in the alkaline solution for 1-4 hours. The membrane was neutralized in 2xSSPE and the DNA photocrosslinked. The membrane was incubated for 1-24 hours in 5 mls of 6xSSPE or SSC, 10xDenhardt's reagent, 0.5% sodium dodecyl sulfate (SDS), and 50-200 µg/ml fragmented and denatured salmon sperm DNA at 42° C.

Southern Blot Hybridization. The location of the *LEU2* gene within the yeast genome was determined by hybridization with the KpnI/EcoRI fragment from pUCLEU2 (12, 28). This fragment was subcloned into pUC19 (KpnI/EcoRI cut) to generate plasmid pUC-KE400. Large quantities of vector free fragment were prepared using the PCR (protocol below) with universal primers UP1200 and UP1201 (Appendix A) and 10 ng of pUC-KE400 DNA template. The PCR fragment (20-100 ng) was random primer radiolabeled with fully degenerate hexanucleotides, α -³²P-dCTP, and the Klenow fragment of DNA polymerase (39, 49). Unincorporated mononucleotides were

eliminated by size exclusion Sephadex G-50 spin chromatography.

Radioactive yields between 15 and 50 million cpm were typically obtained.

The DNA was denatured by the addition of 0.1 volumes of 1 N NaOH with incubation at room temperature for 15 min.

Following prehybridization, the solution was decanted from the membrane and incubated 4-24 hours at 42° C with 4 mls hybridization solution containing 6 x SSPE or SSC, 0.5% SDS, 40% formamide, 50-200 µg/ml fragmented and denatured salmon sperm DNA, and the radiolabeled and denatured PCR fragment from pUC-KE400. After hybridization, the solution was decanted, and the membrane washed twice with 6xSSPE, 0.5% SDS for 10 min. at room temperature, and twice in 1xSSPE, 0.5% SDS for 15 min. at 42° C. The membrane was placed between two pieces of plastic wrap and the radioactive signal detected by autoradiography.

PCR Amplification and Product Digestion. PCR analysis was also used to confirm the presence of the triple helix target site by amplifying all the loci in the genome that were upstream of a *LEU2* gene. If a wild type sequence was present, the PCR product would be sensitive to digestion with *XhoI* and insensitive to the enzyme characteristic of the desired construct. If the wild type sequence had been replaced by recombination, the reverse pattern would be observed. If both wild type and altered sequences were present, PCR products would be partially cut by both enzymes. By selecting for the *LEU2* phenotype, screening recombinants by *LEU2* hybridization, and confirming the presence of the target site by the PCR, several chromosome III constructs containing triple helix target sites were generated.

PCR primers were developed to amplify the region surrounding the *XhoI* site upstream of the *LEU2* gene. The sequence of the *XhoI/SalI* fragment (Fig. 3.7) containing the *LEU2* gene had been fully determined, but the

XmaI
CCCGGGCAGCCTGCTCTGCCTGTGTTTTCTTTAATTGAGCAGTAGACCATTTAGCAGTTGCATGA 65
 PstI
 ATAGCTGCAGATGCAGGGGTGAGTTATGTTGTGCTTACTTTTTAGCTTGCTTGCTGGA
 ACTAC 130
 ATTAGGTATTTGGGGTATTGCTATTGTTACAGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 195
 UP-MID
 TTTTTTTAAAGGCTTACTTACTGATAGTAGATCAACGATCAGTATAATTAAGTCTACAAATGAA 260
 GAGAAATTTAGAAACAGATTTTTTGGCACAAAGGCAAATGAGACTTAGAGATGAAGTATCAGGTA 325
 ATAATTTATACGTATACTACATCGAGACCAAGAAGAACATTGCTGATGTGATGACAAAACCTCTT 390
 CCGATAAAAACATTTAAACTATTA
 AACTAACA
 AATGGATTCCATTAGATCTATTACATTATGGGTGG 455
 A-MID Xho+250
 TATGTTGGAATAAAAATCAACTATCATCTACTA
 ACTAGTATTTACGTTACTAGTATATTATCATA 520
 TACGGTGTTAGAAGATGACGAAAATGATGAGAAATAGTCATCTAAATTAGTGAAGCTGAAACGC 585
 AAGGATTGATAATGTAATAGGATCAATGAATATTAACATATAAAAATGATGATAATAATATTTATA 650
 XhoI
 GAATTGTGTAGAATTGCAGATTCCTTTTATGGATTCC
 TAAATCCTCGACTCGAGGAGA
 ACTTCT 715
 LEU2-Sequence
 AGTATATCCACATACCTAATATTATTGCCTTATTA
 AAAATGGAATCCCAACAATTACATCAAAAT 780
 CCACATTCTCTTCAAAATCAATTGTCCTGTACTTCCTTGTTCATGTGTGTTCAAAAACGTTATAT 845
 Xho-235
 TTATAGGATAATTATACTCTATTTCTCAACAAGTAATTGGTTGTTTGGCCGAGCGGTCTAAGGCC 910
 CCTGATTCAAGAAATATCTTGACC 934

Fig. 3.8. Sequence of pUCLEU2 from the Xho-235 hybridization site to the XmaI ligation site in the pUC19 vector. Oligonucleotides used for sequencing primers or PCR amplification are indicated as arrows within the sequence. The arrowhead indicates the 3' end of the oligonucleotide. Restriction enzyme sites XmaI, PstI, and XhoI are also indicated.

sequence from XhoI to the XmaI ligation site was unknown. This sequence was obtained by dideoxy sequencing of pUCLEU2 using sequencing primers UP1200, LEU2-Sequence, A-MID, UP-MID, and Xho+250 (Fig. 3.8) (Appendix A). The region contained approximately 700 bp of sequence including the small XmaI/PstI fragment that is unlikely to be contiguous on chromosome III.

PCR amplification using primers Xho+250 and Xho-235 (1 μ M each) was conducted on 1.0 μ l of yeast chromosomal DNA embedded in agarose, or 5 μ l of yeast cells lysed by vortexing in a 10 mM EDTA solution containing glass beads. Buffer conditions were 10 mM tris-HCl, pH 8.3 (at 25° C), 50 mM KCl, 1.5 mM MgCl₂, 0.001% (w/v) gelatin, 200 μ M of each dNTP, and 2.5 units of Taq DNA polymerase in 100 μ l total volume. Temperatures and times were 94° C, 1 min.; 55° C, 1 min.; and 72° C, 1 min. for 30 cycles. Reactions were performed under light mineral oil to prevent evaporation. The PCR product (10 μ l) was digested with either XhoI or a construct specific restriction enzyme (A: EcoRI; B: BamHI; C: SpeI; D: HindIII; E: SstI; H: HindIII; P: BamHI) to confirm the presence of the target site in the yeast genome. Products were resolved on a 1.0% agarose gel in 1xTAE and ethidium bromide and visualized by UV illumination.

Yeast constructs SEY6210, A, A₂, B, C, D, and E were prepared and confirmed by this method. In the case of SEY6210B, the hybridization signal on chromosome III was twice as intense as the wild type, and the PCR reactions showed a product partially cut by both XhoI and BamHI. This suggested that there was a duplication of the *LEU2* gene on chromosome III. Subsequent cleavage experiments demonstrated that the duplication flanked the target site resulting in detection of both cleavage products. In the case of SEY6210E, very few transformants were observed and the one which was

subsequently characterized contained an insertion of the target site into yeast chromosome XI. A chromosome III insertion for SEY6210E was not critical, and therefore a second clone was not characterized. Although homologous recombination proceeds with good fidelity, the overall error rate for these constructs likely increased due to the presence of a Ty1 repetitive element upstream of the *LEU2* gene (12, 28).

Affinity Cleaving Reactions. Reactions were performed as described above in footnote 28 (14) and in figure legends.

Enzymatic Cleavage Reactions. Reactions were performed as described above in figure legends 3 and 4 (15) or more completely in the methods section of the *Methods in Enzymology* manuscript (16).

Southern Blotting of Cleavage Products. The 110 and 230 kb cleavage products were detected by hybridization with radiolabeled fragments from *HIS4* (23, 24) and *PGK1* (25, 26), respectively. *HIS4* hybridization was performed with a 250 bp *SalI*/*EcoRI* fragment derived from the *HIS4-SUC2* fusion plasmid YCp503 (24). The *SalI*/*EcoRI* *HIS4* fragment was ligated into pUC19 to generate pUCHIS4-SE250. DNA for hybridization was prepared by the PCR using universal primers UP1200 and UP1201 specific for the pUC19 vector arms. PCR conditions were performed as described above except 10 ng of pUCHIS4-SE250 were used as template. The DNA was radiolabeled and hybridized as described above.

DNA for *PGK1* hybridization was amplified by the PCR directly from pMA91, a plasmid containing the *LEU2* gene, the 2 μ circle, and the *PGK1* gene promoter inserted into pBR322 (25). PCR primers were designed for the amplification of a 1.3 kb fragment from the 5' promoter of *PGK1*. The first primer, PGK.PCR1, was derived from the sequence between the *EcoRI* and *HindIII* sites in pBR322. The second primer, PGK.PCR2, was designed to

hybridize to the promoter region of pMA91 immediately upstream of the transcription initiation site (Appendix A). PGK.PCR2 was designed to avoid amplification of conserved sequences in the promoter to minimize possible cross-hybridization to other loci in the yeast genome. PCR amplification was as described above using 10 ng of pMA91 as template. The PCR product was radiolabeled and hybridized as described above.

References

1. S. A. Strobel, H. E. Moser, P. B. Dervan, Double-Strand Cleavage of Genomic DNA at a Single Site by Triple Helix Formation *J. Amer. Chem. Soc.* **110**, 7927-7929 (1988).
2. G. F. Carle, M. V. Olson, An Electrophoretic Karyotype of Yeast *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3756-3760 (1985).
3. D. C. Schwartz, C. R. Cantor, Separation of Yeast Chromosome-Sized DNAs by Pulsed Field Gradient Gel Electrophoresis *Cell* **37**, 67-75 (1984).
4. G. F. Carle, M. V. Olson, Separation of Chromosomal DNA Molecules from Yeast by Orthogonal-Field-Alteration Gel Electrophoresis *Nuc. Acids Res.* **12**, 5647-5664 (1984).
5. G. F. Carle, M. Frank, M. V. Olson, Electrophoretic Separations of Large DNA Molecules by Periodic Inversion of the Electric Field *Science* **232**, 65-68 (1986).
6. S. M. Clark, E. Lai, B. W. Birren, L. Hood, A Novel Instrument for Separating Large DNA Molecules with Pulsed Homogeneous Electric Fields *Science* **241**, 1203-1205 (1988).
7. B. W. Birren, E. Lai, S. M. Clark, L. Hood, M. I. Simon, Optimized Conditions for Pulsed Field Gel Electrophoretic Separations of DNA *Nuc. Acids Res.* **16**, 7563-7582 (1988).

8. E. Lai, B. W. Birren, S. M. Clark, M. I. Simon, L. Hood, Pulsed Field Gel Electrophoresis *BioTechniques* **7**, 34-42 (1989).
9. R. K. Mortimer, D. Schild, Genetic Map of *Saccharomyces cerevisiae* *Microbiol. Rev.* **49**, 181-212 (1985).
10. T. Satyanarayana, H. E. Umbarger, G. Lindegren, Biosynthesis of Branched-Chain Amino Acids in Yeast: Correlation of Biochemical Blocks and Genetic Lesions in Leucine Auxotrophs *J. Bacteriol.* **96**, 2012-2017 (1968).
11. T. U. Satyanarayana H. E., G. Lindegren, Biosynthesis of Branched-Chain Amino Acids in Yeast: Regulation of Leucine Biosynthesis in Prototrophic and Leucine Auxotrophic Strains *J. Bacteriol.* **96**, 2018-2024 (1968).
12. A. Andreadis, Y. Hsu, G. B. Kohlhaw, P. Schimmel, Nucleotide Sequence of Yeast *LEU2* Shows 5'-Noncoding Region has Sequences Cognate to Leucine *Cell* **31**, 319-325 (1982).
13. J. Dreyer, P. B. Dervan, Sequence-specific Cleavage of Single-stranded DNA: Oligodeoxynucleotide-EDTA•Fe(II) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 968-972 (1985).
14. S. A. Strobel, P. B. Dervan, Site-Specific Cleavage of a Yeast Chromosome by Oligonucleotide-Directed Triple-Helix Formation *Science* **249**, 73-75 (1990).
15. S. A. Strobel, P. B. Dervan, Single-site Enzymatic Cleavage of Yeast Genomic DNA Mediated by Triple Helix Formation *Nature* **350**, 172-174 (1991).
16. S. A. Strobel, P. B. Dervan, Triple Helix Mediated Single-Site Enzymatic Cleavage of Megabase Genomic DNA *Method. Enz.* (In Press).
17. T. J. Povsic, P. B. Dervan, Sequence-Specific Alkylation of Double-Helical DNA by Oligonucleotide-Directed Triple-Helix Formation *J. Amer. Chem. Soc.* **112**, 9428-9430 (1991).
18. T. J. Povsic, S. A. Strobel, P. B. Dervan, Nonenzymatic Double Strand Cleavage of a Yeast Chromosome at a Single Site in High Yield by Triple Helix Formation *J. Amer. Chem. Soc.* (In Preparation).
19. H. E. Moser, P. B. Dervan, Sequence-Specific Cleavage of Double Helical DNA by Triple Helix Formation *Science* **238**, 645-650 (1987).

20. S. A. Strobel, L. A. Doucette-Stamm, L. Riba, D. E. Housman, P. B. Dervan, Site-Specific Cleavage of Human Chromosome 4 Mediated by Triple Helix Formation *Science* **254**, 1639-1642 (1991).
21. C. L. Smith, C. R. Cantor, Preparation and Manipulation of Large DNA Molecules: Advances and Applications *Trend. Biol. Sci.* **12**, 284-287 (1987).
22. C. L. Smith, S. R. Klco, C. R. Cantor, Pulsed-field Gel Electrophoresis and the Technology of Large DNA Molecules in *Genome Analysis: A Practical Approach* K. Davies, Eds. (IRL Press, England, 1988), pp. 41-72.
23. J. K. Keeseey, R. Bigelis, G. R. Fink, The Product of the *his4* Gene Cluster in *Saccharomyces cerevisiae* *J. Biol. Chem.* **254**, 7427-7433 (1979).
24. R. J. Deshaies, R. Schekman, A Yeast Mutant Defective at an Early Stage in Import of Secretory Protein Precursors into the Endoplasmic Reticulum *J. Cell Biol.* **105**, 633-645 (1987).
25. M. J. Dobson, M. F. Tuite, N. A. Roberts, A. J. Kingsman, S. M. Kingsman, Conservation of High Efficiency Promoter Sequences in *Saccharomyces cerevisiae* *Nuc. Acids Res.* **10**, 2625-2637 (1982).
26. M. F. Tuite, M. J. Dobson, N. A. Roberts, R. M. King, D. C. Burke, S. M. Kingsman, A. J. Kingsman, Regulated High Efficiency Expression of Human Interferon-alpha in *Saccharomyces cerevisiae* *EMBO J.* **1**, 603-608 (1982).
27. T. J. Povsic, P. B. Dervan, Triple Helix Formation by Oligonucleotides on DNA Extended to the Physiological pH Range *J. Am. Chem. Soc.* **111**, 3059-3061 (1989).
28. J. R. Broach, J. N. Strathern, J. B. Hicks, Transformation in Yeast: Development of a Hybrid Cloning Vector and Isolation of the *CAN1* Gene. *Gene* **8**, 121-133 (1979).
29. D. A. Fischhoff, R. H. Waterston, M. V. Olson, The Yeast Cloning Vector YEp13 Contains a tRNA₃^{Leu} Gene That Can Mutate to an *Amber* Suppressor *Gene* **27**, 239-251 (1984).
30. J. S. Robinson, D. J. Klionsky, L. M. Banta, S. D. Emr, Protein Sorting in *Saccharomyces cerevisiae*: Isolation of Mutants Defective in the Delivery and Processing of Multiple Vacuolar Hydrolases *Molec. Cell. Biol.* **8**, 4936-4948 (1988).
31. P. B. Dervan, Chemical Methods for the Site-Specific Cleavage of Genomic DNA in *Structure and Methods: Human Genome Initiative &*

DNA Recombination R. H. Sarma, M. H. Sarma, Eds. (Adenine Press, 1990), pp. 37-49.

32. T. J. Povsic, Oligonucleotide Directed Sequence Specific Recognition and Alkylation of Double Helical DNA by Triple Helix Formation (California Institute of Technology, 1992).
33. M. Koob, E. Grimes, W. Szybalski, Conferring Operator Specificity on Restriction Endonucleases *Science* **241**, 1084-1086 (1988).
34. M. Koob, W. Szybalski, Cleaving Yeast and *Escherichia coli* Genomes at a Single Site *Science* **250**, 271-273 (1990).
35. L. J. Maher, B. Wold, P. B. Dervan, Inhibition of DNA Binding Proteins by Oligonucleotide-Directed Triple Helix Formation *Science* **245**, 725-730 (1989).
36. J. C. Hanvey, M. Shimizu, D. Wells, Site-Specific Inhibition of EcoRI Restriction/Modification Enzymes by a DNA Triple Helix *Nuc. Acids Res.* **18**, 157-161 (1990).
37. J. S. Lee, M. L. Woodsworth, L. J. P. Latimer, A. R. Morgan, Poly(pyrimidine)•poly(purine) Synthetic DNAs Containing 5-methylcytosine Form Stable Triplexes at Neutral pH *Nuc. Acids Res.* **12**, 6603-6614 (1984).
38. L. J. Maher, P. B. Dervan, B. Wold, Kinetic Analysis of Oligodeoxyribonucleotide-Directed Triple-Helix Formation on DNA *Biochemistry* **29**, 8820-8826 (1990).
39. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning* (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1989).
40. A. Hinnen, J. B. Hicks, G. R. Fink, Transformation of Yeast *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1929-1933 (1978).
41. H. Ito, Y. Fukuda, K. Murata, A. Kimura, Transformation of Intact Yeast Cells Treated with Alkali Cations *J. Bacterio.* **153**, 163-168 (1983).
42. K. Struhl, D. T. Stinchcomb, S. Scherer, R. W. Davis, High-Frequency Transformation of Yeast: Autonomous Replication of Hybrid DNA Molecules *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1035-1039 (1979).
43. R. P. Moerschell, S. Tsunasawa, F. Sherman, Transformation of Yeast with Synthetic Oligonucleotides *Proc. Natl. Acad. Sci. U.S.A.* **85**, 524-528 (1988).

44. E. M. Southern, Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis *J. Mol. Biol.* **98**, 503-517 (1975).
45. R. K. Saiki, S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, N. Arnheim, Enzymatic Amplification of β -Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia *Science* **230**, 1350-1354 (1985).
46. R. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, H. A. Erlich, Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase *Science* **239**, 487-491 (1988).
47. R. K. Saiki, in *PCR Technology* H. A. Erlich, Eds. (Stockton Press, New York, 1989), pp. 7-16.
48. H. A. Erlich, D. Gelfand, J. J. Sninsky, Recent Advances in the Polymerase Chain Reaction *Science* **252**, 1643-1651 (1991).
49. A. P. Feinberg, B. Vogelstein, A Technique for Radiolabeling DNA Restriction Endonuclease Fragments to High Specific Activity *Anal. Biochem.* **132**, 6-13 (1983).

Chapter IV

Site Specific Cleavage of Human Chromosomal DNA Mediated by Triple Helix Formation

Triple helix experiments on plasmid, bacteriophage λ , and yeast chromosomal DNA demonstrated the feasibility of cutting genomic DNA at a single site within DNA of high complexity (1-4). The specificity and efficiency were sufficient to suggest that triple helix formation could find application in the site specific cleavage of human chromosomal DNA (4). The human haploid genome contains approximately three billion base pairs of duplex DNA (10^9 bp) divided into 23 chromosomes ranging in size from 25-300 Mb. Complete sequence determination of the human genome is a major goal of the human genome initiative (5). A preliminary goal is the identification of genetic markers at approximately 5 centimorgan intervals, or approximately every 5 megabases (Mb) along the length of each chromosome (5, 6). This goal has been reached in some regions of the genome (5-7). In light of this effort, it is desirable to develop a cleavage strategy capable of cutting uniquely at each genetic marker to liberate the intervening megabase fragment for subsequent analysis and subcloning. As an initial demonstration of the site specific cleavage of a human chromosome mediated by oligonucleotide directed triple helix formation, we targeted the tip of human chromosome 4 in a region known to contain the Huntington disease (HD) gene.

Huntington Disease: Review of the Mapping Literature. HD is an autosomal disorder characterized by progressive neurodegeneration (reviewed in 8). Outward characteristics of the disease include involuntary movements by all parts of the body, cognitive deterioration, and often severe

emotional disturbance. The degree of emotional disturbance is variable, but usually includes depression, irritability, apathy, and occasionally violent outbursts, hallucinations and delusions. Patients often develop dysarthria which impairs their speech to the point that they become mute. Patients lose short term memory and organizational abilities, but retain awareness of time, place, and recognition of their own identities and that of family and friends. They are often keenly aware of their intellectual failures.

The biochemical basis of HD is not known, but the disease is characterized physiologically by genetically programmed cell death of specific neuron classes in the caudate nucleus and the basal ganglia (9, 10). Cell loss also occurs in the cortex, globus pallidus, hypothalamus, and cerebellum (8).

Average age of onset is approximately 39 years, but cases of early (ages 2 to 20) and late (ages greater than 60) onset are not uncommon (8). The disease manifests itself at a slow rate, showing symptoms for a period of 3 to 10 years before HD can be definitively diagnosed. The disease has an average duration of approximately 19 years before the death of the patient. Although invariably fatal, HD is the underlying rather than the immediate cause of death. The primary causes of death are aspiration-induced pneumonia, choking, heart failure, hematomas, or suicide.

HD shows a classical pattern of autosomal dominant inheritance at a single locus, but also shows many unusual genetic characteristics. HD has a very low mutation rate. No cases of a new mutation have been documented, suggesting that the mutation arose only once in history and is possibly of European origin (8). A second unusual feature of HD is its inheritance as an autosomal dominant trait that exhibits full penetrance (11). Phenotypic characterization of a number of individuals likely to be homozygous for the mutation demonstrated no increase in the severity of the symptoms or age of

onset compared to heterozygotes for the disease (11, 12). This represents the first human disease that displays complete phenotypic dominance and is suggestive of a gain of function allele (11).

An alternative hypothesis suggests that the HD gene is subject to dominant position-effect variegation (13). In this model a gene located close to a highly condensed heterochromatic region of the chromosome, such as the telomere, may become inactivated by the heterochromatic DNA due to chromosomal rearrangement, insertion, deletion or local region inactivation. The dominant effect comes from trans inactivation of the wild type allele. This pattern of inheritance has been demonstrated in the dominant *Drosophila* mutation *brown*⁺ (14, 15).

A third unusual feature of HD inheritance is the pattern of early onset. Individuals that display early onset typically inherit the HD mutation from the father rather than the mother (16), suggesting that the gene is subject to chromosomal imprinting (17, 18), possibly by an altered methylation state of the inherited DNA (19). It is not clear why paternal inheritance does not always lead to early onset, or why early onset is observed from maternal inheritance.

Efforts to clone the HD gene have relied upon knowledge of its chromosomal location, because the nature of the gene responsible for HD is not known (Reviewed in 20-23). The first major breakthrough in the identification of the chromosomal location for HD came in 1983 (24). Gusella *et al.* identified a restriction length polymorphism (RFLP) in a single copy DNA clone (G8 or D4S10) that showed significant linkage to HD. The clone mapped to chromosome 4 by blotting to a panel of human-mouse somatic cell hybrids. It was the first example of gene localization by RFLP mapping and suggested that the mutation would quickly be identified.

Linkage analysis with additional families confirmed the genetic proximity of D4S10 and HD (25, 26). The maximum recombination frequency between D4S10 and HD was determined by two and three point linkage analysis to be 0.04 (4 cM) (27). As a result of this low, but measurable rate of recombination no linkage disequilibrium was detected between HD and D4S10 (27). Analysis also detected no evidence for locus heterogeneity, an indication that HD is most likely caused by a mutation within a single gene (27).

The location of HD was narrowed further by analyzing chromosomes from patients with Wolf-Hirschhorn syndrome (WHS), a congenital abnormality involving heterozygous deletion of the short arm of chromosome 4 (28). DNA from several of these individuals was assayed for heterozygosity at any of the D4S10 polymorphisms. All of the WHS patients examined displayed homozygosity for each of the RFLP's in D4S10, demonstrating that one copy of D4S10 had been deleted. Analysis of the deletions in these patients placed the HD gene to the terminal band of the short arm of chromosome 4 (4p16) (28). This assignment was confirmed by *in situ* hybridization (29, 30). Multipoint linkage analysis with RFLP markers centromeric to D4S10 demonstrated that the candidate region for the gene was confined to a small segment of DNA (<5 Mb) flanked by D4S10 and the telomere (31, 32). Unfortunately, several years of additional effort has failed to definitively localize HD to a single candidate region within this interval (23).

Extensive effort was invested in isolating genetic markers in the candidate interval between D4S10 and the telomere (33-49). These markers were used to create detailed physical (50) and genetic maps (51) (Fig. 4.1). In an effort to refine the candidate region for HD, RFLP linkage analysis was

performed using several of these markers (51). The design of this work was to identify cross over events between a flanking marker and HD so that the interval could be reduced. It was soon discovered that most of the families that showed recombination between D4S10 and HD failed to narrow the candidate region due to a hot spot for recombination immediately distal to D4S10 (52).

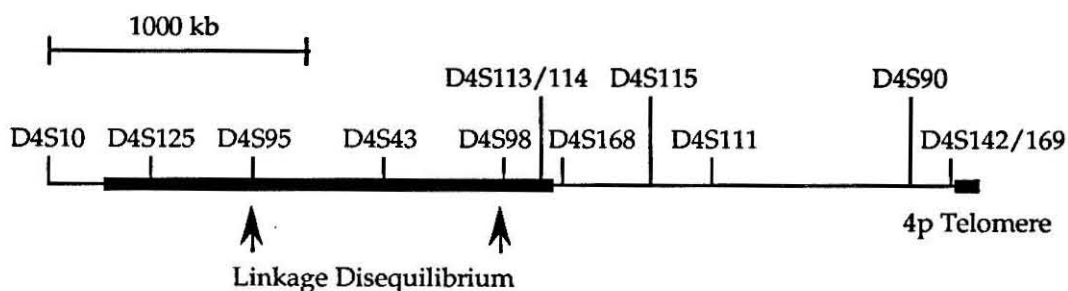


Fig. 4.1. Physical map of interval from D4S10 to 4pter. The genetic markers are indicated along the map by D4S numbers. The thick black lines indicate the two candidate regions for the HD gene based upon genetic linkage analysis. Linkage disequilibrium has been reported between D4S95 and D4S98 and HD, though it was not detected with D4S43. The scale of the map is indicated.

Four recombination events outside the hot spot suggested conflicting locations for HD (51, 53) (Fig. 4.2 and 4.3). In three recombinants, HD segregated with the marker alleles expected for the normal chromosome at D4S10 and at all informative distal markers (Fig. 4.2). The most telomeric informative marker for families 1 and 2 was D4S111, which maps 1100 kb from the telomere of 4p (50) (Fig. 4.2 A, B). Family 3 was also informative for D4S90 which maps to within 300 kb of the telomere (50) (Fig. 4.2 C). Interpretation of these families as single recombinants within the candidate interval placed the HD mutation telomeric to D4S90, within a 300 kb interval

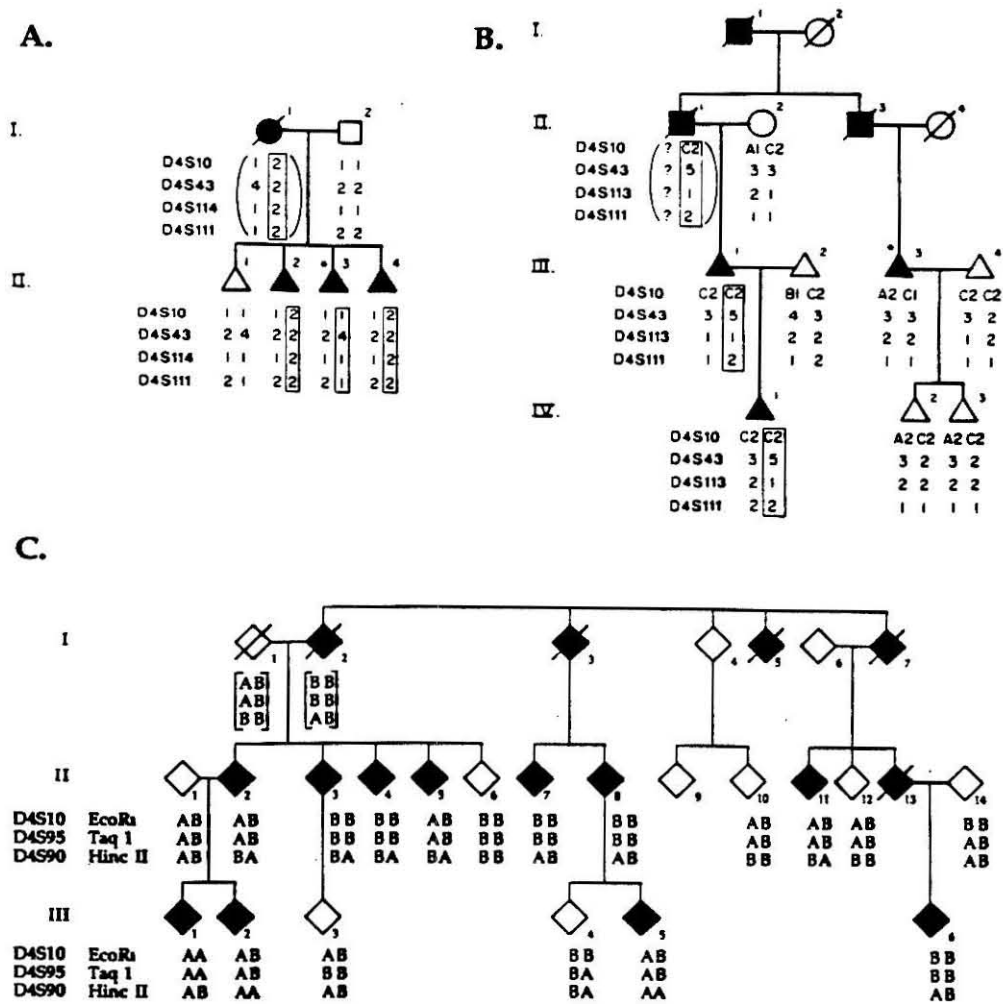
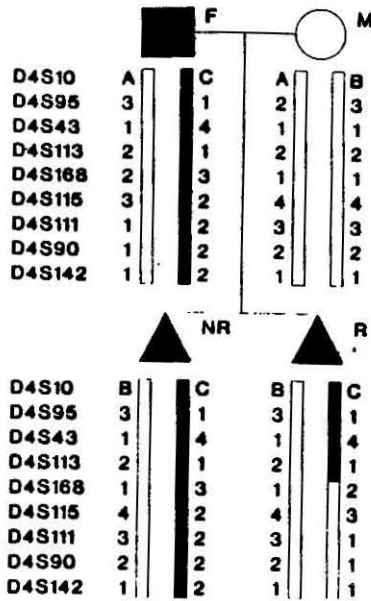


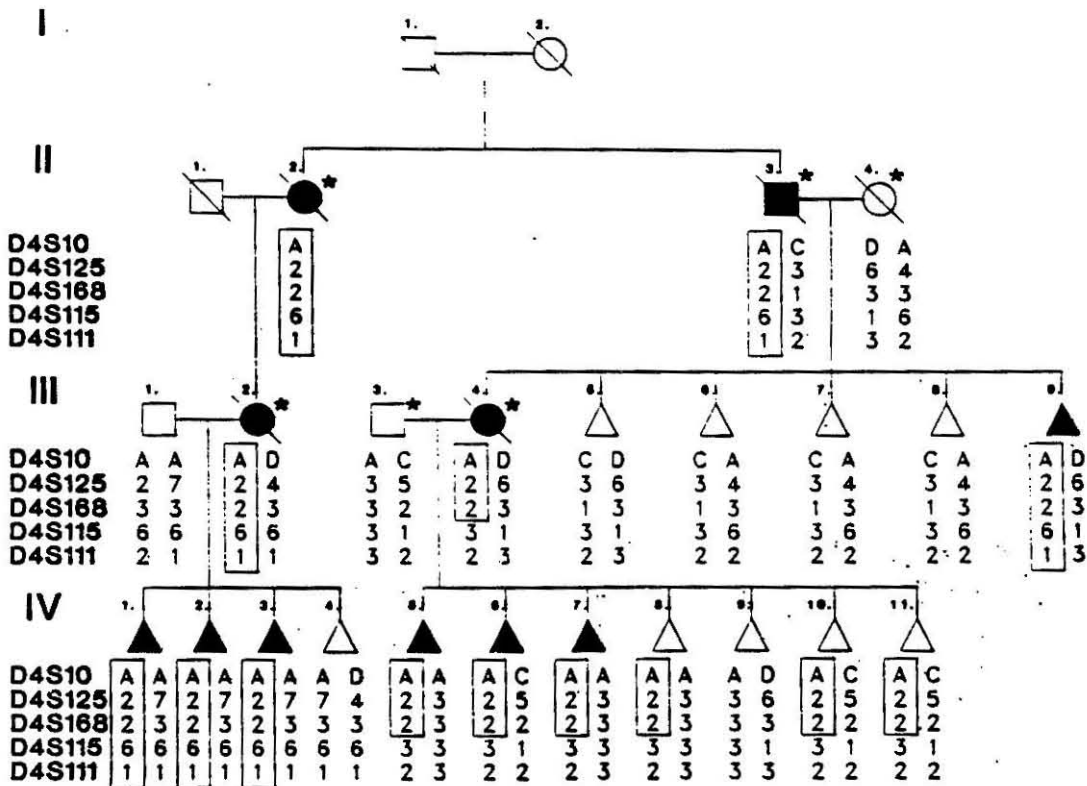
Fig. 4.2. Recombination events predicting a distal location for the HD gene (Taken from references 51 and 53). **A.** Family 1. Filled shapes represented HD affected individuals. The boxed alleles represent the chromosomes carrying the HD gene. The genotype (indicated in brackets) of the affected parent was reconstructed from the progeny shown, and several additional unaffected sibs. Progeny II-3, indicated by an asterisk, reveals a recombination event in which all four marker loci segregate from the HD gene. **B.** Family 2. Graphical representation as in A. Individual III-3, marked by an asterisk, does not possess alleles for any of the marker loci characteristic of the HD chromosome, indicating recombination between the defect and all four DNA loci. Since neither of the children of III-3 are symptomatic, it is not possible to establish which chromosome now carries the HD gene. The presumed recombination event could potentially have occurred during meiosis in individuals I-1, II-1, or II-3. **C.** Graphical representation as in A. HD haplotype in this family is BBA. Affected individual III-1 is a recombinant between the loci D4S10, D4S95, and possibly between D4S90 and the HD gene.

Fig. 4.3. Recombination events predicting an internal location for the HD gene (Taken from references 47, 51 and 61). **A.** Family 4. Affected individuals in the family are depicted as black shapes. Alleles for all informative markers are shown for each chromosome. The solid bar indicates marker alleles characteristic of the HD chromosome in this pedigree. Individual II-2 (marked by an asterisk) reveals a recombination event between D4S113 and D4S115 in which the distal markers segregate from the genetic defect. **B.** Family 5. HD family displaying recombination event between D4S115 and D4S168. The haplotype traveling consistently with HD is boxed. The recombination event occurred during meiosis in individual II-3 and has been passed to affected individuals III-4, IV-5, IV-6, and IV-7. Additional high risk individuals are not yet symptomatic. The telomeric markers segregated from HD in this recombination.

A.



B.



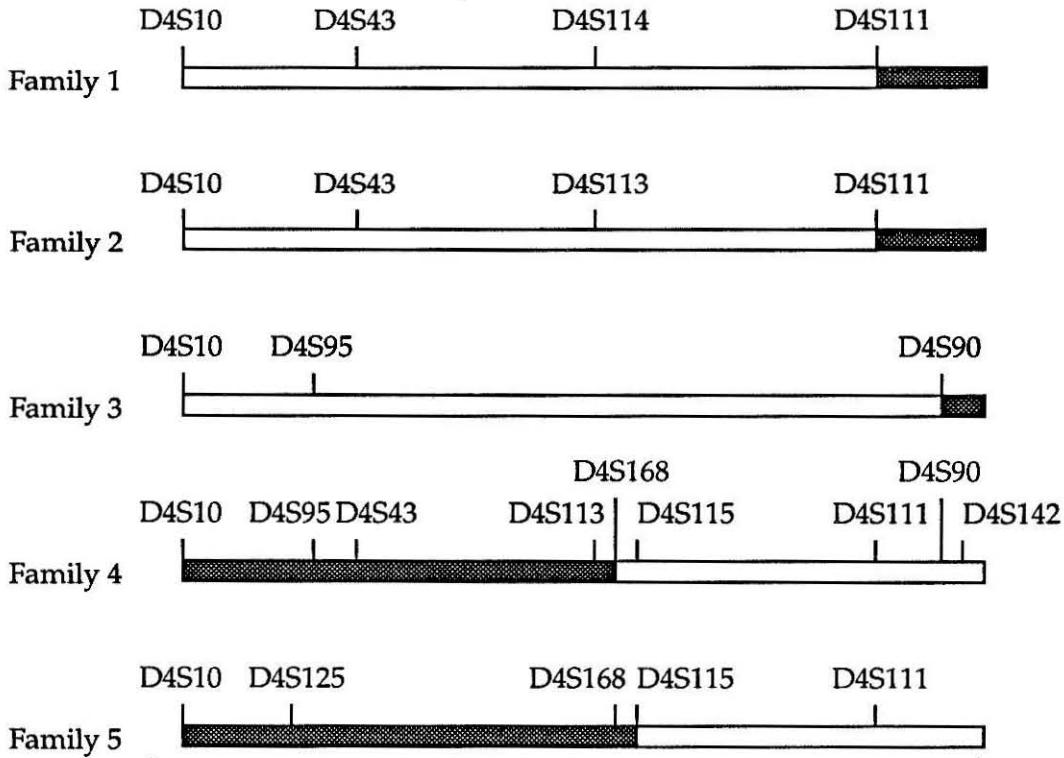


Fig. 4.4. Possible locations for HD based upon five families displaying recombination events informative for this region. Predictions are based on a single recombination event for each recombinant. Hatched boxes correspond to regions whose genotype is unknown. Gray boxes represent regions known or thought to originate from the HD chromosome. Interpretation of each family as a single recombination event predicts two nonoverlapping candidate regions. The conflicting candidate locations could be the result of a double crossover or gene conversion event in a subset of the families.

(Fig. 4.4). However, in family 4, HD segregated with the HD allele at D4S10, and with wild type alleles at telomeric markers D4S111 and D4S90 (51). The most proximal informative marker inherited with the wild type allele was D4S168, which restricts the point of recombination to a 150 kb interval between D4S168 and D4S113 (47) (Fig. 4.3 A). If interpreted as a single recombination event, this family places HD in an interval more than 1 Mb internal from the telomere (50) (Fig. 4.4). This family could be brought into agreement with the other three by postulation of a highly telomeric second recombination event. Using this interpretation of the data, efforts were made to clone the 4p telomere (54, 55).

The telomeric location of the gene was placed in serious doubt when yeast artificial chromosomes containing DNA spanning the region from D4S90 were analyzed (54, 55). The region contained very few genes and no detectable rearrangements associated with the disease. The telomere did not differ in size or structure between HD and normal individuals. Analysis of a polymorphism within 100 kb of the telomere failed to detect recombination in the first three families, and the postulated second recombination event in family 4 was not observed (54). A telomeric location became increasingly doubtful when human chromosomes were shown to contain homologous telomeric sequences which were polymorphic in length, suggestive of telomeric exchange (56). A highly telomeric location might be susceptible to chromosomal exchange, a phenomena inconsistent with tight linkage to 4p16. The final argument against a telomeric location was the observation of linkage disequilibrium between two markers (D4S95 and D4S98) located in the internal candidate region (57-59). Linkage disequilibrium was not observed between HD and the telomeric marker D4S90. This suggests that the

normal haplotypes associated with HD in the first three families might have resulted from a narrow double crossover or gene conversion.

The internal candidate region was further characterized to define the maximum candidate interval for HD (47, 60). Family 4 defined the distal boundary as being between D4S113 and D4S168 (47). Only a single family had been reported to carry a recombination event between D4S10 and HD that was distal to D4S95 but proximal to D4S90 (44). Such a recombinant would define the lower boundary for the internal candidate region. Unfortunately, the recombinant had been originally mistyped for D4S95 and the recombination was actually in the D4S10 recombination hot spot. An additional recombination event was recently identified that supports, but does not narrow, the assignment of HD to the internal candidate region (61) (Fig. 4.3 B, Fig. 4.4). Thus, the boundaries for the internal HD candidate region remain D4S10 and D4S168, a 2.5 Mb interval containing no known recombinants (60).

A complete analysis of linkage disequilibrium in the candidate region demonstrated the intractability of the HD mapping problem (61). Whereas cloning of the cystic fibrosis gene was assisted by identification of significant linkage disequilibrium (62), no continuous increase in linkage disequilibrium was detected across the HD candidate region, nor was a region of extreme disequilibrium identified (61). While disequilibrium was detected at D4S98 and D4S95 using a subset of RFLP's, disequilibrium was not consistently observed with closely spaced RFLP's in these loci (61).

The diversity of HD haplotypes suggest that simple recombination following a single original HD mutation cannot readily explain the modern pool of HD chromosomes (61). Possible explanations for the complex pattern of disequilibrium in the candidate region are: (i) none of the markers are sufficiently close to HD to display extreme disequilibrium, (ii) two or more

original and independent HD mutations in different primordial haplotypes now represent a significant portion of the HD gene pool, and/or (iii) a single HD mutation has been transferred to a different background haplotype by a high rate of double recombination or gene conversion (61).

The failure to define the HD locus by genetic linkage or disequilibrium dictates the need for alternative strategies capable of identifying physical alterations in the DNA. These efforts now include saturation of the candidate region with potential transcripts, searches for physical rearrangements, and the plans for large-scale genomic sequencing. Isolation of DNA from the candidate regions of HD affected and normal chromosomes could provide a means to physically identify the HD mutation. The well characterized nature of the 4p16.3 region makes it an ideal test case for enzymatic cleavage of human genomic DNA mediated by triple helix formation.

This chapter describes the site specific cleavage of human chromosome 4 at the genetic marker D4S10 to liberate a 3.6 Mb fragment containing both the proximal and distal candidate regions. It outlines several procedures for screening unsequenced genetic markers for triple helix target sites overlapping methylase/endonuclease sequences. It describes the application of these strategies to the identification of a target site within D4S10 and the subsequent characterization and cleavage of that site within plasmid, yeast, and human genomic DNA. Finally, it describes efforts to make a flanking cut on chromosome 4 at a marker likely to be distal to the HD gene. These techniques are readily generalizable to any genetic marker for the orchestrated cleavage of the human genome into megabase sized fragments readily manipulated by standard procedures of molecular biology.

Part I

Identification of Triple Helix Target Sites Within Unsequenced Genetic Markers

The insertion of a triple helix target site by genetic engineering or homologous recombination was a relatively straightforward procedure when targeting plasmid, bacteriophage λ or yeast genomic DNA (1-3). Target site insertion into a predetermined genetic locus in a human chromosome is sufficiently difficult to make it highly undesirable. Selection for recombinants is complex and insertion is infrequent due to the low rate of recombination in mammalian cells (63, 64). Additionally, it would be necessary to insert a target site into each of two genetic markers for cleavage to liberate a product resolvable by pulsed field gel electrophoresis. Considering the difficulty of target site construction, the ability to target endogenous sequences in genomic DNA would be very advantageous. The triple helix motif is sufficiently generalizable that it should obviate the need for target site insertion, however some sequence information is necessary to specifically target sites in genetic markers.

Although a large number of genetic markers are available for each human chromosome, the DNA sequence generally remains undetermined. Typically a short 500-1000 bp segment has been sequenced for use as a sequence tagged site (STS) using the polymerase chain reaction (PCR). The time necessary to completely sequence large segments of cloned genomic DNA has been steadily reduced, yet it would be desirable to screen cloned genetic markers for homopurine target sites that overlap methylase-endonuclease sequences in a directed manner with minimal sequencing effort.

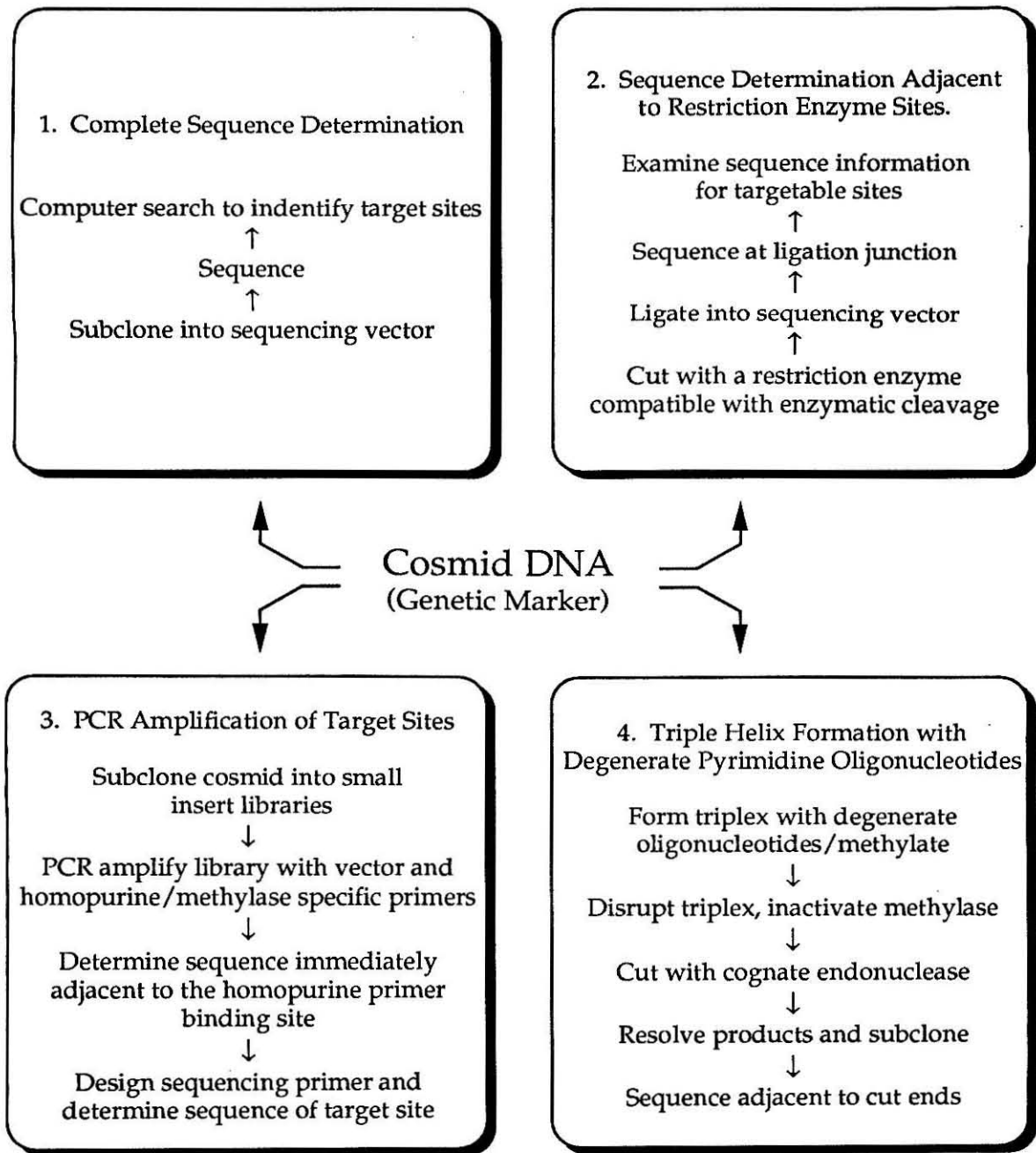


Fig. 4.5. Strategies for detecting triple helix binding sites that overlap methylase-restriction enzyme sequences.

Strategies for Detection of Target Sites. A number strategies to identify triple helix binding sites that overlap a methylase recognition site could be imagined (Fig. 4.5), including: (i) complete sequence determination and target site identification by a computer search, (ii) partial sequencing of DNA from a restriction enzyme library to determine the sequences immediately adjacent to all the sites of a given restriction enzyme, (iii) PCR amplification using partially degenerate primers specific to homopurine sequences overlapping a restriction enzyme site, and/or (iv) enzymatic cleavage of cosmid DNA mediated by triple helix formation with partially degenerate pyrimidine oligonucleotides. Each strategy has its particular strengths and weaknesses, but each has proven useful for identifying target sites for enzymatic cleavage mediated by triple helix formation.

Complete sequence determination is easily the most laborious technique for target site identification, but the resulting sequence information can be rigorously searched for the binding sequences of all triple helix motifs and any overlap of these sequences with a methylase recognition site. A computer search can be conducted to identify sites with mismatches, crossovers, poly-G runs, etc. adjacent to any methylation site, both those commercially available and those that would require purification. Although this is the least imaginative strategy and the most labor intensive, it would identify all target sites within a genetic marker.

A more selective sequencing effort involves the creation of a library for a specific restriction enzyme. All the sequences immediately adjacent to that restriction enzyme site could be determined by sequencing from vector specific primers. If a homopurine target site was not identified among the sequences for a particular enzyme, a library could be constructed with a second restriction enzyme, and the process repeated. Although this is a

straightforward procedure capable of identifying a wide variety of target sites, it is limited by the number of libraries that are constructed and fully analyzed, and the inability to conveniently make a continuous sequence determination on both sides of each restriction site. This reduces the utility of alternate strand triple helix formation when examining the sequence data. Despite these limitations, this strategy successfully identified two target sites flanking the Ewings sarcoma locus on chromosome 22 (65).

The PCR assay is a rapid technique for the detection of target sites (66, 67). In this approach at least two libraries are constructed from the genetic marker using unrelated restriction enzymes. Pooled plasmid DNA purified from the transformation culture is used as a template for the PCR using a vector primer and a partially degenerate primer for amplification. The degenerate primer is designed with a 3' end specific for the sequence of the methylase being targeted, an internal degenerate segment to hybridize to a wide variety of homopurine sequences, and a 5' anchor to improve amplification yield and serve as a primer site for subsequent sequence determination. Analysis with this system can be used to detect overlap with any sequence of interest, including those recognized by methylases not commercially available. It can also be used to detect homopurine sequences targetable by affinity cleaving that do not overlap a methylation site.

Although target detection is rapid, identification of the target sequence is more difficult. The PCR product can not be directly used for target site sequence determination because the target site is directly hybridized by the degenerate oligonucleotide resulting in loss of the original sequence information during amplification. Instead, the target sequence can be determined by sequencing the internal portion of the amplification product by priming at the anchor, synthesizing a second primer complementary to a

sequence distal to the degenerate region, and sequencing back to the target site on the original cosmid clone.

A procedure of high fidelity that enables rapid analysis of a number of enzyme systems utilizes partially degenerate oligonucleotides for triple helix formation followed by enzymatic methylation and cleavage (4, 68-70). The oligonucleotides contain two domains, a pyrimidine degenerate domain designed to bind to a wide variety of homopurine and mixed sequences, and a nondegenerate domain designed to overlap a restriction enzyme site. If a cleavage product is observed after methylation and restriction enzyme digestion, it indicates that a site bound by a fraction of the degenerate oligonucleotide mixture overlapped an enzyme recognition site. The sequence can be identified by subcloning the cleavage products, and sequencing immediately adjacent to the ligation site. The technique is limited to searches with methylases that are readily available. This procedure was used to identify the target site within D4S10 on human chromosome 4.

Partial Sequence Determination of D4S10. D4S10 was the first genetic marker linked to a human disease by RFLP analysis (24). It maps to 4p16.3 and is approximately 4 Mb from the telomere (50). It is known to be proximal to the Huntington disease locus, but does not show linkage disequilibrium with the disease (24, 27). Cosmid clone 8C10I5, a 38 kb insert into the 8.1 kb pWE15 cosmid vector, contains the majority of the D4S10 locus, and was used for target site identification. Three STS's were previously identified within this region for RFLP analysis (71), but a computer search of this sequence (approximately 600 bp) did not identify a target site.

The cosmid was partially sequenced by shotgun cloning to generate approximately 5.0 kb of noncontinuous sequence information. A computer search identified a number of homopurine target sites that were subsequently

targeted by affinity cleaving with oligonucleotides containing EDTA•Fe(II), but the background cleavage was so extensive that all the chromosomes were degraded into megabase sized fragments and further studies of affinity cleaving on human chromosomal DNA were not pursued. No triple helix target sites that overlapped a methylase/endonuclease site were identified within the sequence.

Although sequencing of 8C10I5 could have been pursued more extensively, it became evident that a more directed approach to target site identification was needed. Two strategies were employed for target site identification within 8C10I5: (i) PCR amplification using partially degenerate oligonucleotides as primers and a pool of 8C10I5 subclones as template, and (ii) enzymatic cleavage mediated by partially degenerate pyrimidine oligonucleotides for triple helix formation. The fidelity of each technique was initially tested on the pUCLEU2 series of plasmids.

Detection of Triple Helix Target Sites by PCR. Partially degenerate oligonucleotide primers were synthesized containing three domains (Fig. 4.6). The primers were designed to detect triple helix sites overlapping TaqI, AluI, HaeIII, dam, EcoRI, and MspI methylation sites. An additional primer specific for general homopurine sequences was also designed. For restriction enzymes with 5'-pyrimidine half sites (TaqI, MspI and the general homopurine oligonucleotide), the second domain was degenerate in C and T, while for enzymes with 5'-purine half sites (AluI, HaeIII, dam, and EcoRI) the second domain was degenerate in A and G.

As an initial demonstration of the ability of this procedure to identify the desired class of sequences, PCR amplification was performed on plasmids pUCLEU2, pUCLEU2A, and pUCLEU2B. The later two plasmids each contain a homopurine sequence at a defined location, but have different restriction

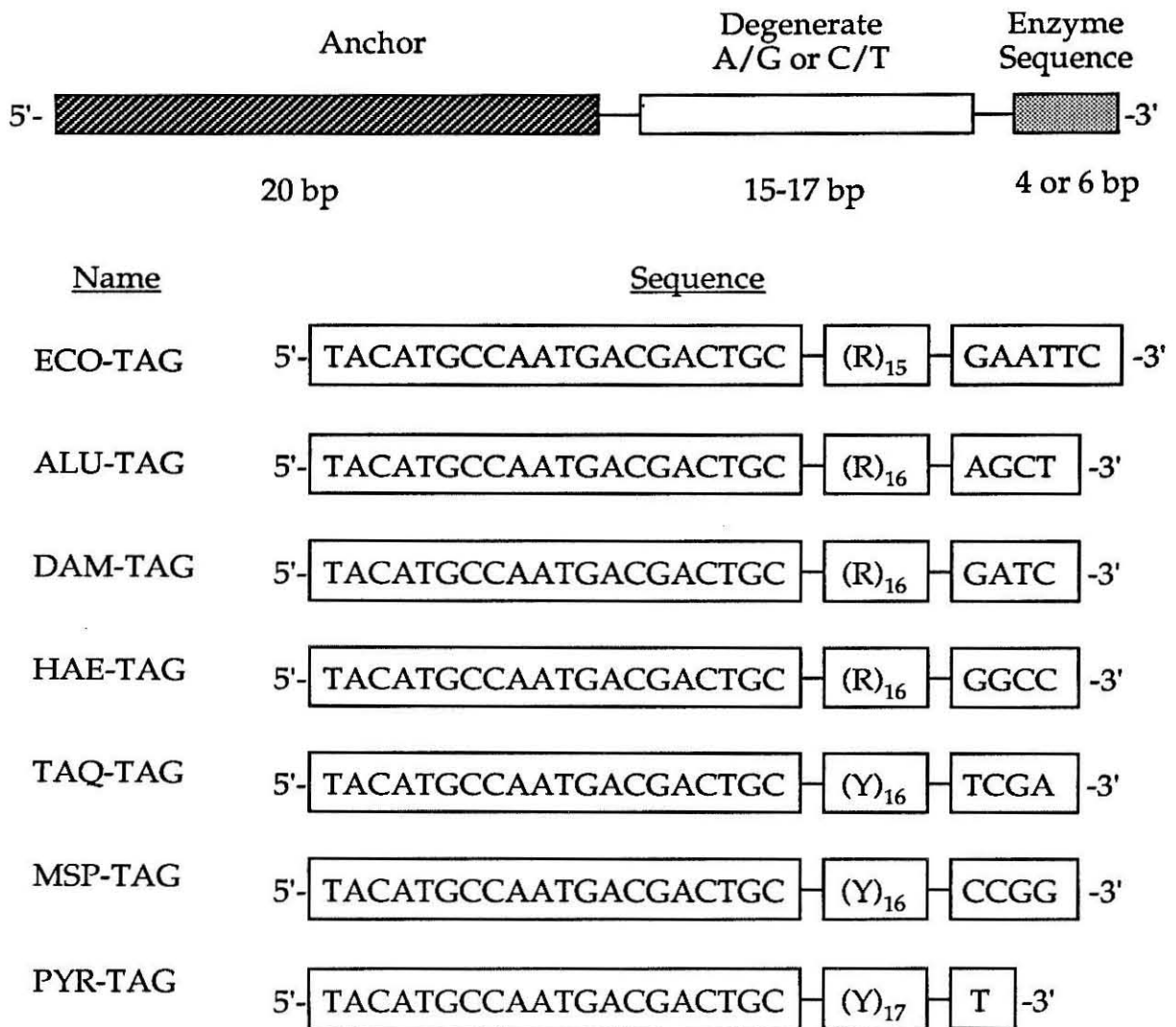


Fig. 4.6. Sequences of oligonucleotide primers used for the detection of homopurine sequences overlapping restriction enzyme sites by PCR. The primers contain three domains: (i) a 3' sequence complementary to the restriction enzyme (gray), (ii) a purine or pyrimidine degenerate domain (white), and (iii) a sequence tag at the 5' end for subsequent sequence identification (strip). R represents 50% each of A and G, and Y represents 50% each of C and T.

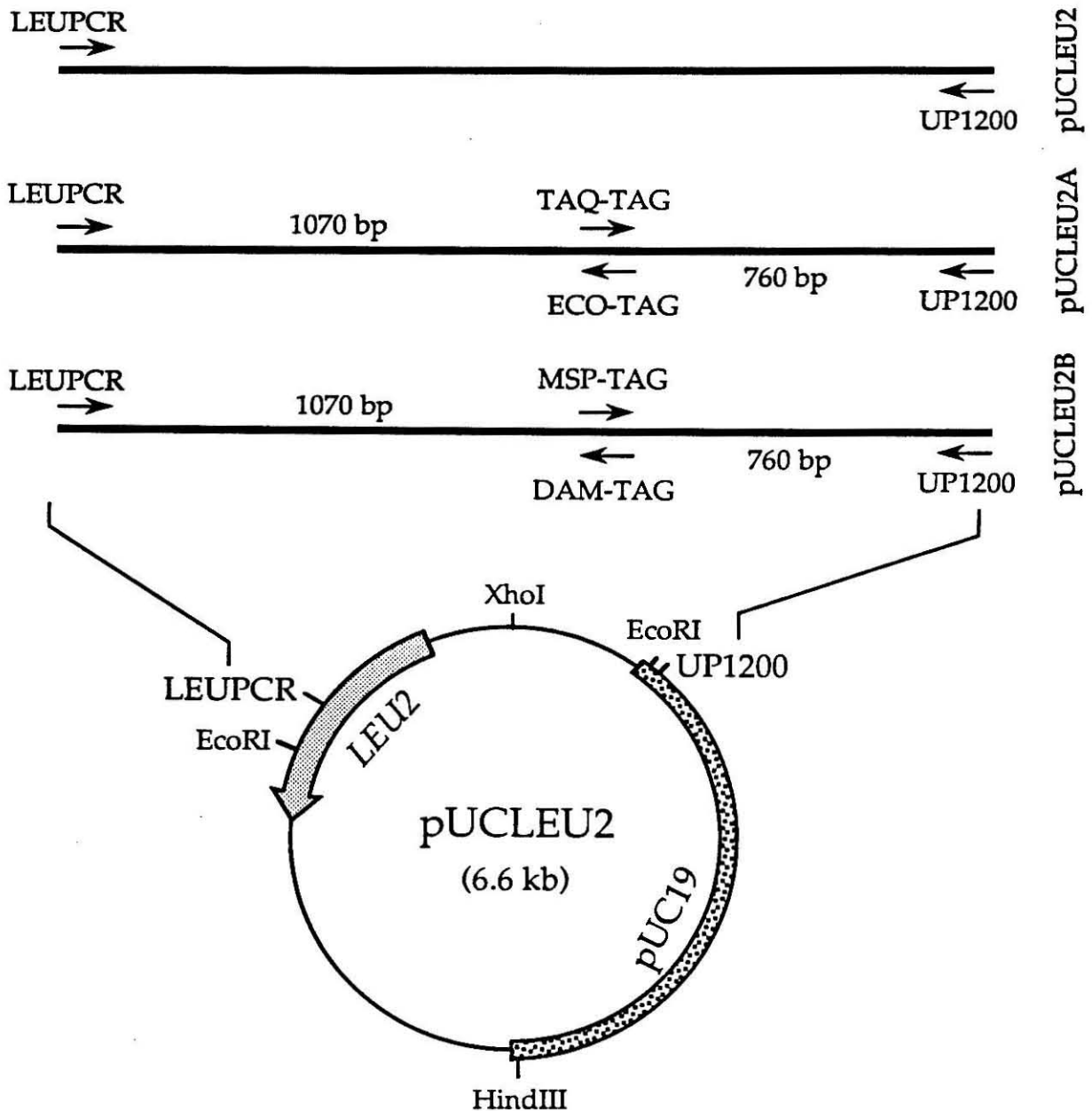


Fig. 4.7. Orientation of the PCR primers at the predicted hybridization sites on the pUCLEU2 plasmid series. The point of the arrow below each primer indicates the 3' end of the oligonucleotide. The distance between the primer hybridization sites is indicated in base pairs. Primer PYR-TAG (not shown) is oriented in the same direction as TAQ-TAG and MSP-TAG and is predicted to amplify a 760 bp product from both pUCLEU2A and B. Primers ALU-TAG and HAE-TAG (not shown) are not predicted to have amplification products within the region of the inserted duplex. The primer sequences are indicated in Fig. 4.3 and in Appendix A.

enzyme sites at the 5' and 3' ends of the oligonucleotide binding site (Fig. 4.7). The plasmids served as ideal substrates for determining the ability of the assay to correctly distinguish general homopurine sites from the narrower class of sites targetable by enzymatic cleavage. Because of the orientation of the homopurine sequences in the pUCLEU2 plasmids, the second primer used for purine degenerate oligonucleotides was LEUPCR, a primer from the *LEU2* gene with the 3' end oriented toward the *XhoI* site (Fig. 4.7). Amplification with pyrimidine degenerate primers was performed with UP1200, a pUC19 universal primer that also had its 3' end oriented toward the *XhoI* site (Fig. 4.7). In this trial study, sequence assignments were not determined from the amplification products. Instead, the size of the product was used to determine the site of primer hybridization and the sequence assigned from the known sequence of pUCLEU2. If degenerate oligonucleotide hybridization to the *XhoI* duplex insert lead to amplification, priming by LEUPCR was expected to generate a 1070 bp product, while priming by UP1200 should produce a 780 bp fragment (Fig. 4.7).

PCR amplification using the *EcoRI* specific oligonucleotide, ECO-TAG, and LEUPCR as primers with 1 ng of either of the three plasmids as template, demonstrated the strength of this technique (Fig. 4.8 A). No amplification was detected in the absence of template or in the presence of either the parent plasmid or a plasmid with a *BamHI* site instead of *EcoRI* to the 5' end of the homopurine sequence (lanes 1, 2 and 4). In the presence of pUCLEU2A, a plasmid with a 24 bp homopurine sequence overlapping an *EcoRI* site, a single 1070 bp fragment was detected in good yield (lane 3). No products of other sizes were observed suggesting that the reaction had high specificity for the desired sequence.

Fig. 4.8. PCR amplification of the pUCLEU2 plasmid series using a partially degenerate primer specific for a single methylase recognition sequence and either UP1200 or LEUPCR primers. All amplification reactions were performed in 10 mM tris-HCl, pH 8.3 (at 25° C), 50 mM KCl, 1.5 mM MgCl₂, 0.001% (w/v) gelatin, 200 μM of each dNTP, 2 μM in each oligonucleotide primer, and 2.5 units of Taq DNA polymerase in 100 μl total volume using 1 ng of plasmid DNA as template. Temperatures cycles were 94° C 1.0 min., 2.0 min. at an annealing temperature dependent upon degenerate primer, and 72° C for 3.0 min. for 30 cycles. **A.** Amplification with ECO-TAG and LEUPCR with annealing temperature of 65° C. **B.** Amplification with TAQ-TAG and UP1200 at annealing temperature of 65° C. **C.** Amplification with DAM-TAG and LEUPCR annealed at 60° C. **D.** Amplification with MSP-TAG and UP1200 with annealing temperature of 65° C. **E.** Amplification with ALU-TAG and LEUPCR with annealing temperature of 65° C. **F.** Amplification with HAE-TAG and LEUPCR annealed at 65° C. **G and H.** Amplification with PYR-TAG and UP1200 at annealing temperatures of 60° C and 40° C, respectively.

A. EcoRI

Primers	{	ECO-TAG	+	+	+	+
		LEUPCR	+	+	+	+
Templates	{	pUCLEU2	-	+	-	-
		pUCLEU2A	-	-	+	-
		pUCLEU2B	-	-	-	+
			1	2	3	4

1070 bp →



B. TaqI

Primers	{	TAQ-TAG	+	+	+	+
		UP1200	+	+	+	+
Templates	{	pUCLEU2	-	+	-	-
		pUCLEU2A	-	-	+	-
		pUCLEU2B	-	-	-	+
			5	6	7	8

780 bp →



C. dam

Primers	{	DAM-TAG	+	+	+
		LEUPCR	+	+	+
Templates	{	pUCLEU2	+	-	-
		pUCLEU2A	-	+	-
		pUCLEU2B	-	-	+
			9	10	11

1070 bp →



D. MspI

Primers	{	MSP-TAG	+	+	+	+
		UP1200	+	+	+	+
Templates	{	pUCLEU2	-	+	-	-
		pUCLEU2A	-	-	+	-
		pUCLEU2B	-	-	-	+
			12	13	14	15

780 bp →



E. AluI

Primers	{	ALU-TAG	+	+	+	+
		LEUPCR	+	+	+	+
Templates	{	pUCLEU2	-	+	-	-
		pUCLEU2A	-	-	+	-
		pUCLEU2B	-	-	-	+
			16	17	18	19



F. HaeIII

Primers	{	HAE-TAG	+	+	+	+
		LEUPCR	+	+	+	+
Templates	{	pUCLEU2	-	+	-	-
		pUCLEU2A	-	-	+	-
		pUCLEU2B	-	-	-	+
			20	21	22	23

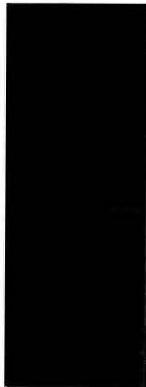
650 bp →



G. Pyrimidine (60° C)

Primers	{	PYR-TAG	+	+	+
		UP1200	+	+	+
Templates	{	pUCLEU2	+	-	-
		pUCLEU2A	-	+	-
		pUCLEU2B	-	-	+
			24	25	26

780 bp →



H. Pyrimidine (40° C)

			+	+	+
			+	+	+
			+	-	-
			-	+	-
			-	-	+
			27	28	29

← 2700 bp

← 1200 bp

← 780 bp



High fidelity amplification was also detected with other primers, but the specificity of amplification for some primers was imperfect. A pattern similar to EcoRI was detected with a TaqI specific primer (Fig. 4.8 B). A 780 bp amplification product was only detected from plasmid pUCLEU2A (lane 7). The dam primer showed a major 1070 bp amplification product when pUCLEU2B was used as template (Fig. 4.8 C, lane 11), but with either of the other plasmids three minor products were detected with sizes of 400, 650, and 2000 bp (lane 9, 10). Unfortunately, none of these minor sites contained a complete dam sequence.

The MspI degenerate primer was the least specific of the primers tested (Fig. 4.8 D). It amplified the insert site from pUCLEU2B (lane 15), but also weakly amplified pUCLEU2A (lane 14). Plasmid pUCLEU2B contained an MspI site (5'-CCGG-3') overlapping a homopurine sequence, but the homopurine sequence in pUCLEU2A contained a 5' end that was significantly different than an MspI site (5'-TCGA-3'). This raised serious concerns about the fidelity of the MspI primers for target site identification.

No designed target site was present for the AluI and HaeIII primers. As expected, no significant amplification was detected with AluI using LEUPCR as the second primer (Fig. 4.8 E). The HaeIII primer also showed no amplification at the duplex insertion site, but interestingly it showed a major 650 bp product in all three plasmids (Fig. 4.8 F, lanes 21, 22, and 23). The sequence 630 bp from the LEUPCR primer (20 bp of sequence was added by the anchor region in the degenerate primer) contained a complete HaeIII site overlapping a 18 bp homopurine sequence interrupted by a single T (5'-AAAAGGAAAGGTGAGAGGCC-3'). Two other minor amplification products were also detected, but neither of these products corresponded to a sequence with a complete HaeIII site.

PCR reactions were also performed using a primer specific for general homopurine sequences, i.e., no methylation site was specified. Amplification at 60° C showed a 780 bp product with both the pUCLEU2A and B plasmid templates, consistent with general amplification of homopurine sequences (Fig. 4.8 G, lanes 25 and 26). Reducing the annealing temperature to 40° C resulted in amplification of two other products, 1200 and 2700 bp in size, in all three plasmids (Fig. 4.8 H, lanes 27-29). A sequence search approximately 1200 and 2700 bp from the UP1200 primer revealed an extensive homopurine sequence at each map position. The sequences 5'-AGAAAAAGGAA-AGGTGAGAG-3' and 5'-AAGAAGGAGAAAAAGGAGGA-3' were detected 1250 and 2696 bp from UP1200, respectively. This was a highly encouraging result, as it rapidly identified the major homopurine sequences near the nondegenerate primer.

While the general level of fidelity was good for each primer, some amplification of incorrect sequences was detected. Although easily disregarded in this study, incorrect amplification would severely complicate identification of target sites if an unsequenced template were used for amplification. Considering the effort required to identify the homopurine sequence from amplification products, this low level of improper amplification could make the technique counterproductive. Additionally, the high temperature of oligonucleotide hybridization necessary to achieve good fidelity of sequence differentiation would favor sequences with a high GC content. Because such sequences are not readily targeted by triple helix formation, the assay could miss the sites best suited for cleavage. Nevertheless, experiments were conducted to identify a target site within D4S10 using this technique.

Detection of a Homopurine Target Site in D4S10 using PCR

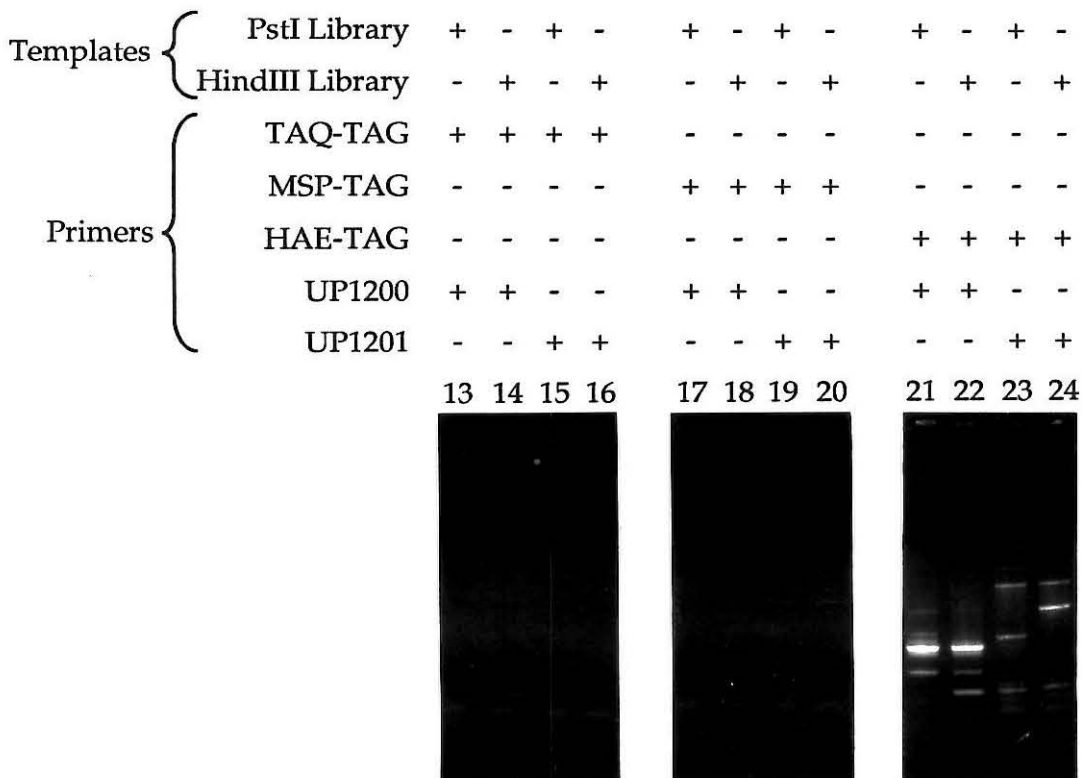
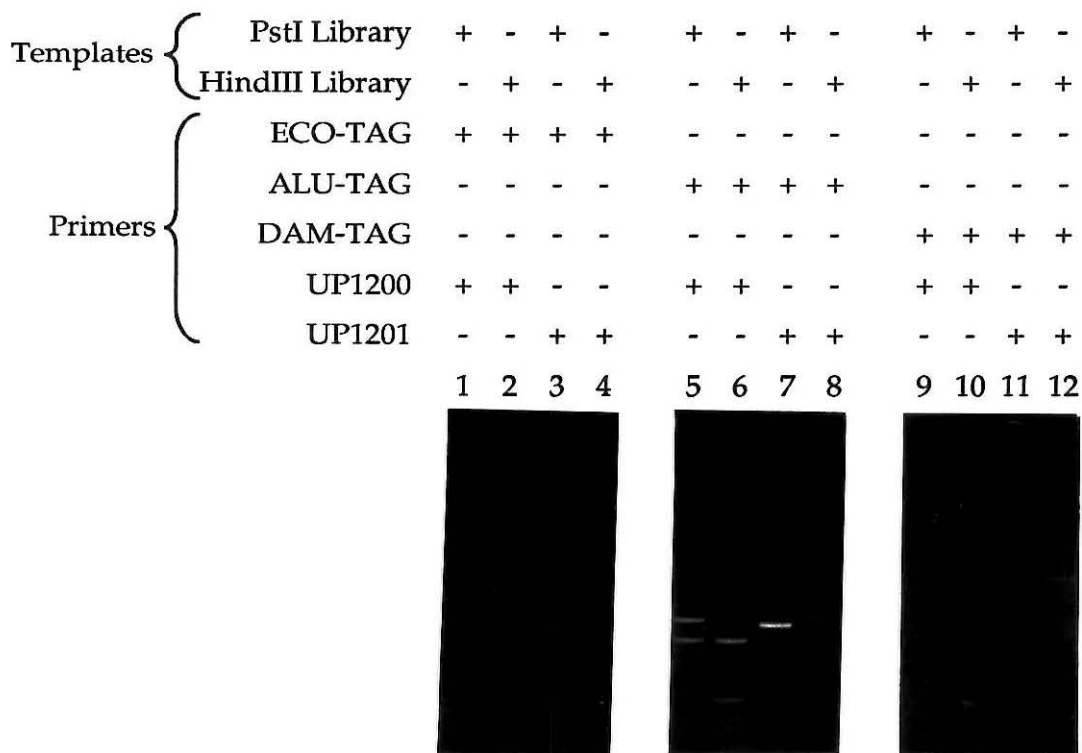
Amplification. The PCR was used to search cosmid clone 8C10I5 for target sites. Six libraries of 8C10I5 were constructed by single and double digestion with HindIII, PstI, and EcoRI, followed by ligation into pUC19 linearized with the appropriate enzyme(s), and transformation to recover the correct ligation products. Total plasmid DNA from the transformation culture was isolated and used as template in the PCR reaction.

Amplification reactions were individually performed using degenerate primers specific for six restriction enzymes, and either of the universal primers from the pUC19 vector arms (UP1200 or UP1201). Reactions were performed on both the PstI and HindIII libraries to improve coverage of the cosmid in case a target site should be too distant from a ligation site or incorrectly oriented within a given restriction fragment. Bands common to both libraries facilitated the identification of products likely resulting from amplification of the recircularized vector.

No amplification was detected using EcoRI, dam, TaqI, and HpaII specific primers (Fig. 4.9). However, at least one discrete amplification product was detected in each reaction with the AluI and HaeIII primers (lanes 5-8 and 21-24). AluI amplification had proceeded with good fidelity in the trial studies suggesting that the products should be the result of amplification from a target site. The fidelity of the HaeIII primer was suspect in trial studies, and though the amplification was encouraging, it had to be considered with some skepticism. Amplification of different sized products was also detected in the other four libraries, consistent with the presence of a target site for both AluI and HaeIII.

Attempts to identify the sequence recognized by the degenerate oligonucleotides were unsuccessful. This included direct sequencing of the

Fig. 4.9. Amplification of total 8C10I5 library DNA with a universal primer and a partially degenerate oligonucleotide specific for the recognition sequence of a single methylase. All amplification reactions were performed in 10 mM tris-HCl, pH 8.3 (at 25° C), 50 mM KCl, 1.5 mM MgCl₂, 0.001% (w/v) gelatin, 200 μM of each dNTP, 2 μM in each oligonucleotide primer, and 2.5 units of Taq DNA polymerase in 100 μl total volume using 10 ng of plasmid DNA as template. Temperatures cycles were 94° C 1.0 min., 60° C 2.0 min., and 72° C for 3.0 min. for 30 cycles. Appearance of amplification products suggestive of the presence of a homopurine sequence overlapping the methylation site.



purified PCR products (72) and sequencing from templates prepared by asymmetric PCR (73). No interpretable sequence information was obtained from these experiments. It should be possible, however, to access the sequence from the PCR products if additional effort was invested in PCR sequencing.

Detection of Target Sites by Triple Helix Formation with Degenerate Oligonucleotides. Potential infidelity of the PCR assay combined with the difficulties involved in determining the sequence of the target site indicated that development of another procedure would be valuable. A possible strategy of potentially high fidelity involved enzymatic cleavage mediated by degenerate pyrimidine oligonucleotide directed triple helix formation (Fig. 4.10). If a cleavage product was observed in this technique, it represented a site protectable by triple helix formation and cleavable by the restriction enzyme, the two essential criterion of the target site. Although the cleavage yield was not expected to be high using degenerate oligonucleotides, the products could be directly subcloned to determine the sequence of the target site.

Partially degenerate oligonucleotides were designed with two domains: a nondegenerate domain two to three bases in length specific for the sequence recognized by a single methylase, and a 16 base degenerate domain specific for a wide class of homopurine and mixed sequences (Fig. 4.11). Nucleotide analogs and degenerate ratios were carefully chosen to optimize target site recognition and cleavage efficiency. MeC and BrU base analogs were chosen because MeC substituted oligonucleotides have higher binding affinities and BrU substitution results in oligonucleotides with reduced binding specificity (3, 4, 74).

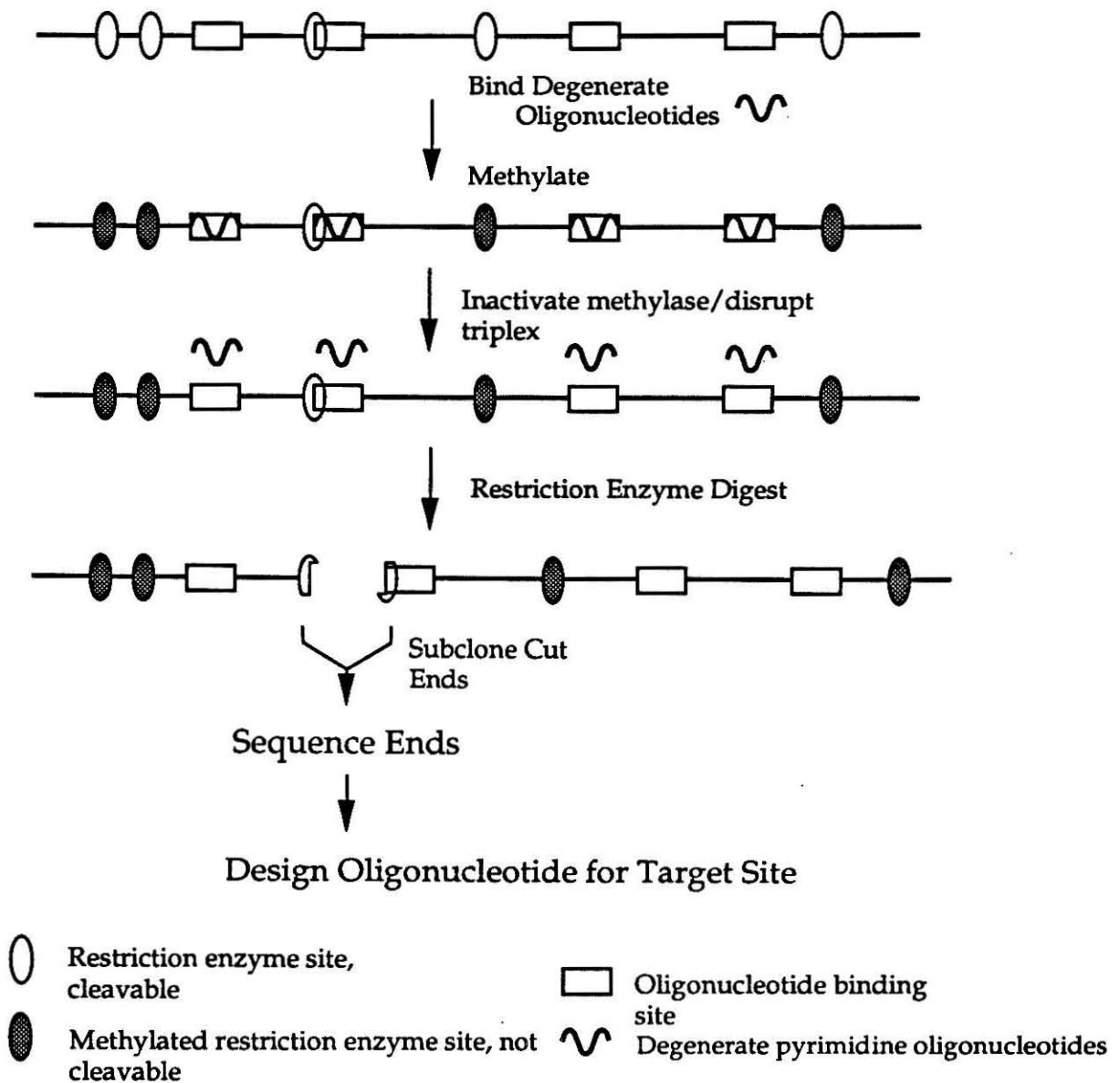
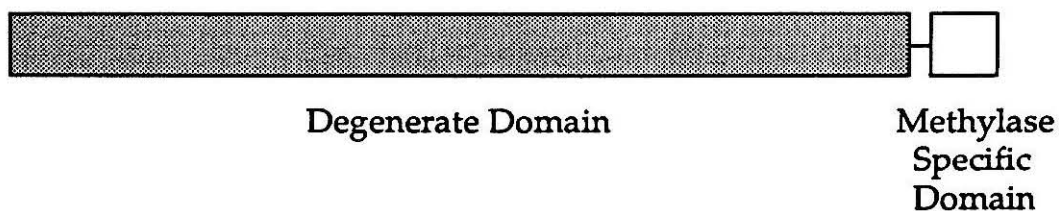


Fig. 4.10. General scheme for identification of sequences targetable by triple helix mediated enzymatic cleavage. Linearized cosmid DNA is equilibrated with a partially degenerate oligonucleotide containing a nondegenerate end specific for a single methylation site. The DNA is methylated to completion, modifying the DNA such that all sites not bound by the oligonucleotide are resistant to subsequent restriction enzyme digestion. The methylase is removed by phenol extraction, the DNA ethanol precipitated, and the triple helix disrupted by resuspending in a high pH restriction enzyme buffer. The appearance of a cleavage product upon restriction enzyme digestion demonstrated that a site bound by a fraction of the oligonucleotide mixture overlaps a methylation site. Subcloning of the cut ends and sequencing of the resulting ligation junction identifies the sequence of the target site. Using this sequence information and the general rules of oligonucleotide directed recognition of duplex DNA, an oligonucleotide specific for the target site can be designed. The process is repeated with several different enzyme combinations until a cleavage product is detected and a target site identified.



<u>Oligonucleotide Sequence</u>	<u>Oligonucleotide Name</u>	<u>Methylase Sequence</u>
5'- YYYYYYYYYYYYYYYYYY T ^{MeC} -3'	Alu-Me	5'-AGCT-3'
5'- YYYYYYYYYYYYYYYYYY MeC T -3'	dam-Me	5'-GATC-3'
5'- YYYYYYYYYYYYYYYYYY MeC T T -3'	Eco-Me	5'-GAATTC-3'
5'- YYYYYYYYYYYYYYYYYY MeC MeC -3'	Hae-Me	5'-GGCC-3'
5'- MeC MeC YYYYYYYYYYYYYYYYYY -3'	Msp-Me	5'-CCGG-3'
5'- MeC T YYYYYYYYYYYYYYYYYY -3'	Taq-Me	5'-TCGA-3'

Y = 60% BrU and 40% MeC

Fig. 4.11. (Above) Schematic of the degenerate and methylase specific domains comprising the oligonucleotides used in the enzymatic sequence search. (Below) Sequence listings of oligonucleotides specific for various methylases. Y represents a nucleotide ratio of 60% BrU and 40% MeC at the indicated base positions.

Sequences with high G content or contiguous G tracts are not efficiently targeted by the pyrimidine triple helix motif near neutral pH (1, 74). To minimize the concentration of oligonucleotides in the population specific for such sequences, a nucleotide ratio of 60% BrU and 40% MeC was used at each degenerate position. This was done to reduce the concentration of oligonucleotides in the mixture that were incapable of binding. The result was a pool of 65,536 unique sequences unequally represented in the population (Fig. 4.12 A). At this nucleotide ratio, 85% of the molecules in the pool had a BrU/MeC base ratio of 50% or higher. At 10 μ M total oligonucleotide, the (BrU)₁₆ sequence had the highest concentration at 3 nM, and the (MeC)₁₆ sequence had the lowest at 4 pM. The oligonucleotide composition with the highest total concentration were sequences with an overall base content of (BrU)₁₀(MeC)₆ at 2 μ M or 0.25 nM in each sequence (Fig. 4.12 B).

The detection of imperfectly matched homopurine or mixed sequences was further improved by performing the triplex and methylation reactions at slightly acidic pH (1-4). This increased the fraction of oligonucleotides that were capable of binding a target site, bringing the effective binding concentration of the oligonucleotide to a level where a cleavage signal could be detected.

Reactions were initially performed using a partially degenerate oligonucleotide specific for EcoRI methylase on plasmid pUCLEU2A. While it was known that triple helix formation can mediate cleavage at the pUCLEU2A target site, it was not known if detectable cleavage could be achieved using a degenerate oligonucleotide. Linearized plasmid pUCLEU2A DNA (3 nM) was incubated with a partially degenerate oligonucleotide (10 μ M) specific for EcoRI in the presence of polycation to facilitate triple helix

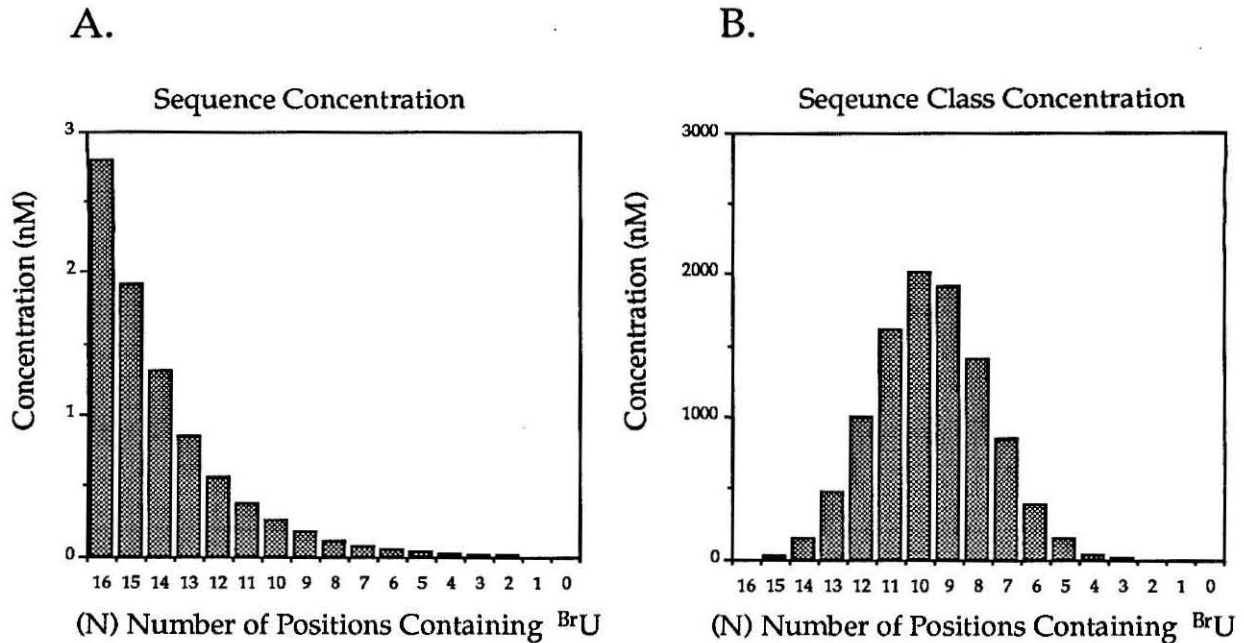


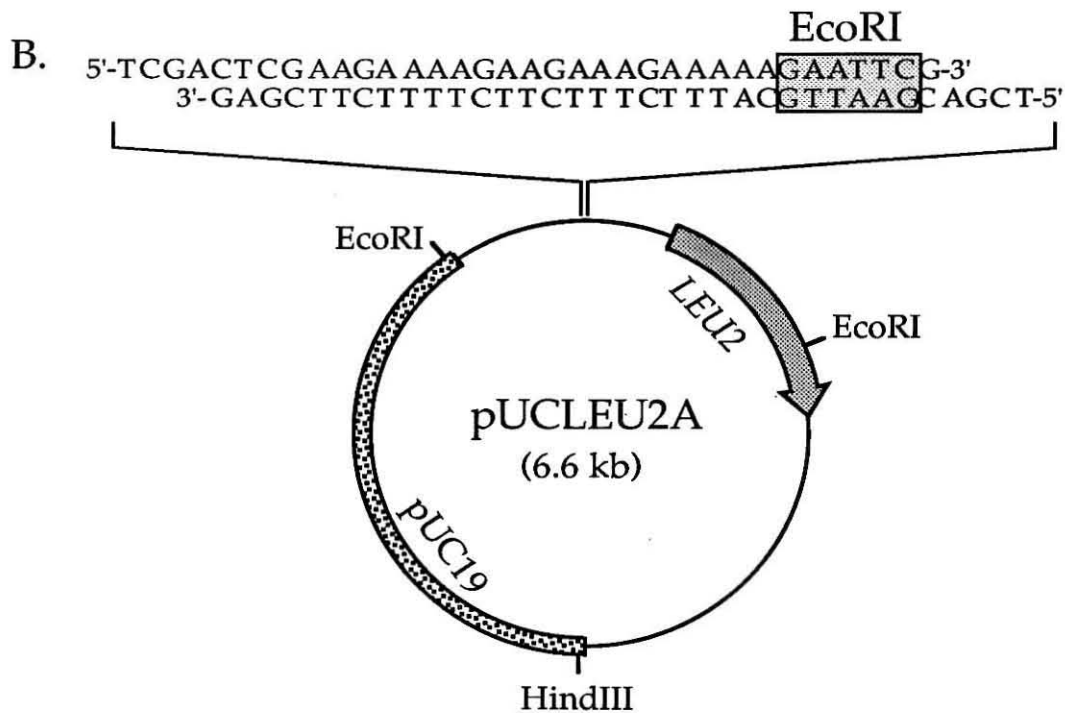
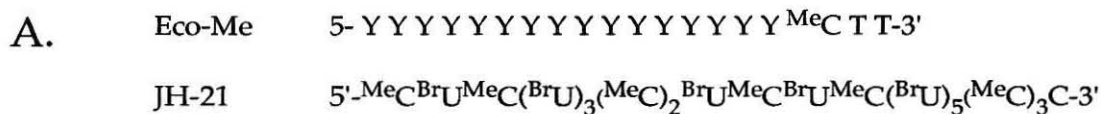
Fig. 4.12. A. Oligonucleotide concentration of individual sequences as a function of base composition. An equal ratio of MeC to BrU in the phosphoramidite mixture would have resulted in an equal concentration of all oligonucleotides in the solution. The 60% BrU and 40% MeC ratio resulted in higher concentrations for oligonucleotides predominantly containing BrU. The (BrU)₁₆ oligonucleotide was most concentrated, and the (MeC)₁₆ oligonucleotide was the least concentrated. **B.** Oligonucleotide concentrations of the general sequence class (BrU)_N(MeC)_{16-N} as a function of base composition. An equal ratio of MeC to BrU in the phosphoramidite mixture would have generated a concentration distribution with a maxima at (BrU)₈(MeC)₈. This would have meant that more than half the oligonucleotide concentration would have had a base composition of 50% MeC or greater. The 60/40 ratio used in the synthesis of these oligonucleotides resulted in a concentration distribution in which 85% of the oligonucleotides have a base composition of 50% BrU or greater. Because G rich target sites are not readily bound by pyrimidine oligonucleotides, this ratio increases the oligonucleotide concentration capable of binding potential target sites.

formation. The DNA was methylated to completion with EcoRI methylase and digested with EcoRI restriction enzyme following methylase inactivation and triplex disruption. Reactions containing a nondegenerate MeCBrU oligonucleotide with no specificity for the target site were also performed as a control against general methylase inhibition.

The degenerate oligonucleotide was able to mediate the site specific cleavage of pUCLEU2A at the desired target site with an efficiency approaching 20% at pH 6.6 (Fig. 4.13, lane 1). The efficiency increased inversely with the pH of the reaction mixture, as expected considering the pH sensitivity of triple helix formation (lanes 1-3). Very little cleavage was observed at pH 7.4 (lane 3). Unmethylated DNA was cut to completion indicating that the triplex had been efficiently disrupted (lane 4). No cleavage was detected using the nonspecific oligonucleotide at any pH (lanes 5-7), confirming that the cleavage was not due to nonspecific methylase inhibition by the modified oligonucleotide.

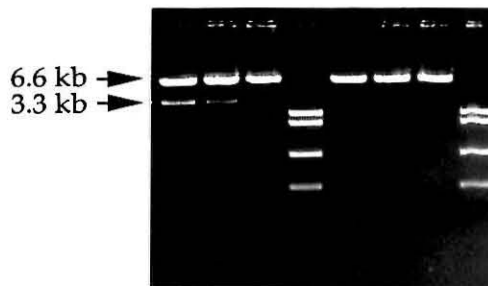
The relative ratios of target site to oligonucleotide suggest that the oligonucleotide modifications and low pH were essential to detection of this site. At 10 μ M total oligonucleotide, the concentration specific for the pUCLEU2A target site was 0.8 nM. This would produce a maximum cleavage yield of 28% if every oligonucleotide were bound to its target site. However, an oligonucleotide specific for this target site showed a pH stability up to pH 7.8, and 0.8 nM was about a factor of ten below the dissociation concentration reported for triple helix formation to a similar sequence (75). The fact that little cleavage was seen even at pH 7.4 suggested that the target site was being protected by oligonucleotides with only partial sequence identity, and that this was essential for efficient signal detection.

Fig. 4.13. Site specific cleavage of plasmid pUCLEU2A using a partially degenerate oligonucleotide specific for EcoRI methylase. **A.** Sequence of oligonucleotides Eco-Me and JH-21. Y indicates a ratio of 60% BrU and 40% MeC at each indicated position. **B.** Map of plasmid pUCLEU2A showing the HindIII site used to linearize the plasmid, and the three EcoRI sites. The sequence of the target site is shown as an insert above the plasmid. It contains an EcoRI site overlapping the 24 bp homopurine sequence. **C.** Enzymatic cleavage of plasmid DNA at the target site mediated Eco-Me oligonucleotide. Plasmid pUCLEU2A (250 ng) was linearized with HindIII and incubated with 10 μ M oligonucleotide Eco-Me in 2 mM spermine, 100 mM NaCl, 100 mM tris-HCl, pH 6.6-7.4, 10 mM EDTA for 12-14 hours at 22° C. S-adenosyl methionine (160 μ M), BSA (100 μ g/ml) and EcoRI methylase (40 units) were added to the reaction (final volume 20 μ l) and incubated at 22° for 9.5 hours. The DNA was phenol extracted, ethanol precipitated, 70% ethanol washed, dried, and resuspended in EcoRI restriction enzyme buffer containing 100 mM potassium glutamate, 25 mM tris-acetate pH 9.5, 10 mM magnesium acetate, 50 μ g/ml BSA, 1 mM 2-mercaptoethanol, and 20 units EcoRI restriction enzyme. Digested the DNA to completion at 37° C for one hour. Resolved products on 0.8% 1xTBE agarose gel containing ethidium bromide. Visualized products by UV excitation. The intact linearized plasmid DNA is 6.6 kb in size, while the desired products are each 3.3 kb fragments.



C.

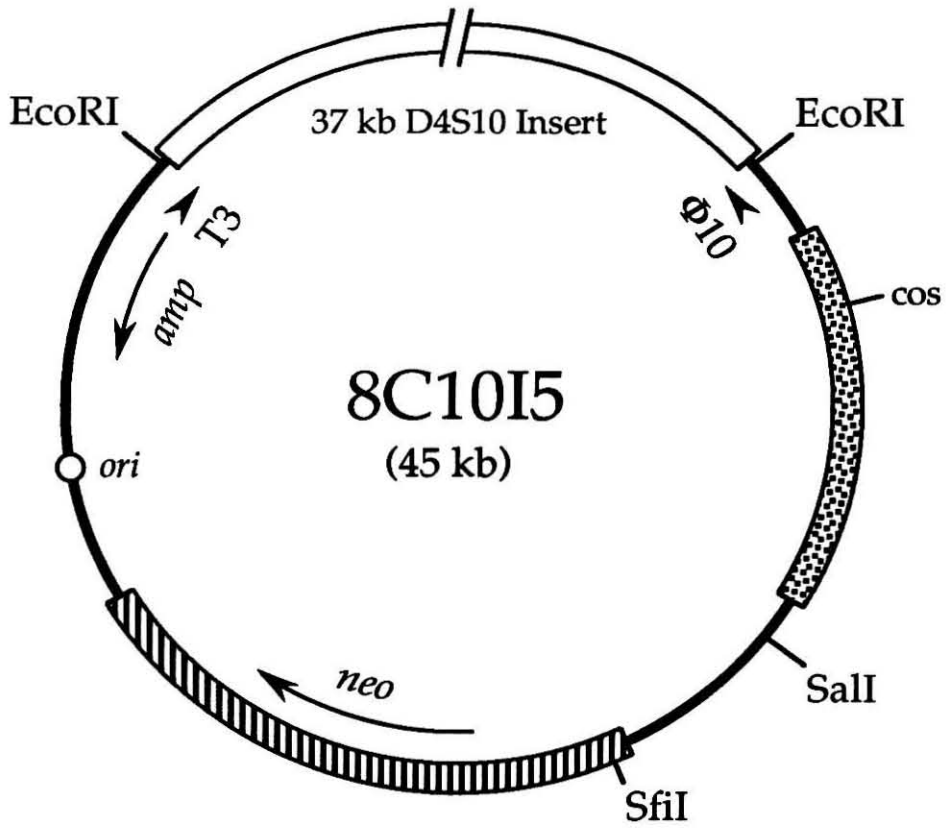
	pH	6.6	7.0	7.4	6.6	6.6	7.0	7.4	6.6
Oligonucleotide Eco-Me		+	+	+	+	-	-	-	-
Oligonucleotide JH-21		-	-	-	-	+	+	+	+
EcoRI Methylase		+	+	+	-	+	+	+	-
EcoRI Endonuclease		+	+	+	+	+	+	+	+
		1	2	3	4	5	6	7	8



PCR amplification with degenerate primers detected an endogenous homopurine sequence overlapping a HaeIII site within pUCLEU2. To roughly determine the range of sequences detectable with degenerate oligonucleotides for triple helix formation, a partially degenerate oligonucleotide specific for HaeIII was incubated with linearized pUCLEU2 plasmid DNA. No cleavage was detected under any of the pH's tested, including pH 6.6. Failure to detect this site was likely caused by the high G content of the sequence, particularly at the 3' end near the HaeIII site. In addition to the two G's in the HaeIII site, six of the next nine bases were G, and one of the remaining three was a T. This sequence constitutes a good example of the need for several rapid strategies for detecting target sites, as each technique will have target sequences that are difficult to identify.

Detection of a Target Site in D4S10 by Enzymatic Cleavage. Cosmid clone 8C10I5 from genetic marker D4S10 was searched for a target site by enzymatic digestion mediated by degenerate oligonucleotide directed triple helix formation. Cosmid DNA was linearized with the single cutting enzyme SallI, and incubated with one of six different partially degenerate oligonucleotides to promote triple helix formation. The DNA was modified with the appropriate methylase and worked up as indicated (Fig. 4.14). Enzymatic treatment with EcoRI, HaeIII, HhaI, HpaII, and TaqI failed to detect any cleavage products. PCR had suggested that a HaeIII site might be present, but none was detected using this assay. Treatment with AluI, however, revealed 38 and 7 kb cleavage products cut in 4% yield (Fig. 4.14, lane 3). Considering the central position of the SallI site within the pWE15 vector and the size of the cleavage products, the cleavage site had to be within the D4S10 insert. Reactions performed on DNA linearized with SfiI, a second enzyme that only cuts once in the vector, showed that the 7 kb cleavage product

Fig. 4.14. Detection of a target site in 8C10I5 by AluI methylation and digestion mediated by a partially degenerate oligonucleotide specific to AluI (Alu-Me). **Upper.** Map of cosmid 8C10I5 containing pBR322 (solid line), bacteriophage λ DNA (speckled line), SV40 and neomycin resistance gene (stripped line), and D4S10 insert DNA (white line, not drawn to scale). Also shown are the SalI and SfiI restriction sites used for cosmid linearization and mapping of the target site. The cosmid contains approximately 37 kb of insert DNA from D4S10 ligated into unique BamHI site of the cosmid vector pWE15. The vector contains selection for ampicillin and neomycin resistance. **Lower.** Enzymatic cleavage of 8C10I5 DNA using a partially degenerate oligonucleotide specific for AluI methylase. 8C10I5 cosmid DNA (250 ng) was linearized with SalI and incubated with 10 μ M oligonucleotide Alu-Me in 1 mM spermine, 50 mM NaCl, 50 mM tris-HCl, pH 6.6, 10 mM EDTA, and 1 mM 2-mercaptoethanol for 12-14 hours at 22° C. S-adenosyl methionine (160 μ M), BSA (100 μ g/ml) and AluI methylase (10 units) were added to the reaction (final volume 20 μ l) and incubated at 22° for 4.0 hours. The DNA was phenol extracted, ethanol precipitated, 70% ethanol washed, briefly dried, and resuspended in AluI restriction enzyme buffer containing 10 mM Bis Tris Propane-HCl, 10 mM MgCl₂ 1 mM dithiothreitol, pH 9.5 (25° C) and 8 units AluI restriction enzyme. Digested the DNA to completion at 37° C for one hour. Resolved products on 1% agarose pulsed field gel in 0.5x TBE, 2.0 sec. switch times, 120° switch angle, 6 V/cm, at 14° C for 10 hours. Cleavage products were detected by ethidium bromide staining and UV excitation. To quantitate the cleavage efficiency and improve the detection signal for reproduction, the gel was transferred to Nytran membrane and hybridized with total 8C10I5 radiolabeled DNA.



Oligonucleotide Alu-Me - - +

AluI Methylase - + +

AluI Endonuclease - + +

1 2 3

45 kb -
38 kb -

7 kb -



became smaller, indicating that the target site was less than 2 kb distal to the T3 promoter of the pWE15 vector (Fig. 4.14) (76, 77).

Sequence Identification of the Target Site by Subcloning. The identity of the target site was determined by subcloning the 7 kb cleavage product and sequencing at the ligation site. While the low cleavage efficiency made it difficult to obtain sufficient material for subcloning, it was only necessary to subclone one of the products for complete sequence determination. The 7 kb product retained the majority of the pBR322 vector used in the construction of pWE15, including the origin of replication and the ampicillin resistance gene (Fig. 4.15). For this reason, the most straight forward approach to subcloning this fragment was to recircularize the DNA and select for ampicillin resistance upon transformation. AluI restriction enzyme cuts to leave blunt ends. The Sall site contains a four base overhang with a recessed 3' end. The DNA was recircularized by filling in the Sall site with the Klenow fragment of DNA polymerase and ligating to the blunt ended AluI site. This ligation regenerated the Sall site, so the resulting transformants were checked for proper ligation by cutting with Sall, and by digestion with SfiI (Fig. 4.15). A single colony was identified that contained the properly sized products.

The target site was transferred to a sequencing vector by cutting the recircularized product with Sall and EcoRI, an enzyme that cut once within the recircularized plasmid at the opposite boundary of the pWE15 vector and the D4S10 insert (Fig. 4.15). This product was ligated into pUC19 and the sequence adjacent to the Sall site determined by dideoxy sequencing from the universal primer UP1201. The resulting sequence contained the expected AluI half site partially contained within the Sall site used for ligation. The AluI site overlapped the 19 bp homopurine sequence 5'-AATGGA-AAGAGAGAGAGAG-3' at the 3' end. The complete sequence on both sides

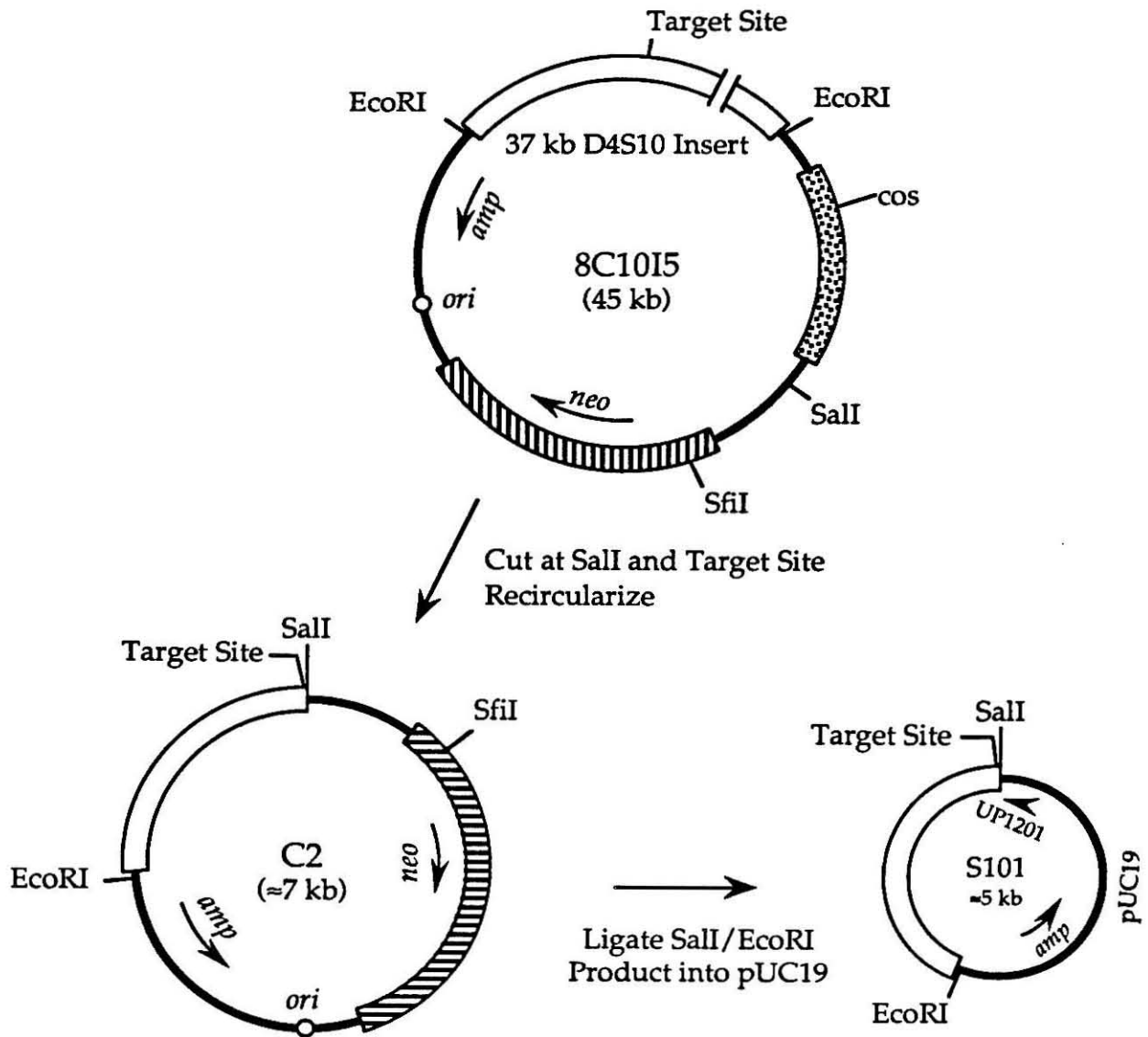


Fig. 4.15. Subcloning and sequence determination of the AluI/triple helix target site in cosmid clone 8C10I5. The open arches correspond to DNA from genetic marker D4S10 subcloned into cosmid vector pWE15. The striped region was derived from SV40 viral DNA into which the neomycin-resistance gene from transposon Tn5 was inserted. The segment containing small squares was derived from bacteriophage λ and contains the *cos* site. The solid black line is pBR322 plasmid DNA. The origin of replication and the ampicillin resistance gene from pBR322 are indicated. Total cosmid DNA was linearized with SalI and cut at the AluI/triple helix site as described in Fig. 4.9. The resulting 7 kb fragment was gel purified and recircularized with T4 DNA ligase. Transformation and selection with ampicillin yielded plasmid C2 which contained the triple helix site adjacent to the SalI ligation junction. The SalI/EcoRI fragment of C2 was ligated into pUC19 for sequencing. The target site was adjacent to the SalI site and was identified by dideoxy sequencing using UP1201 as primer. Sequence on both sides of the AluI site were identified by sequencing 8C10I5 cosmid DNA from an upstream primer.

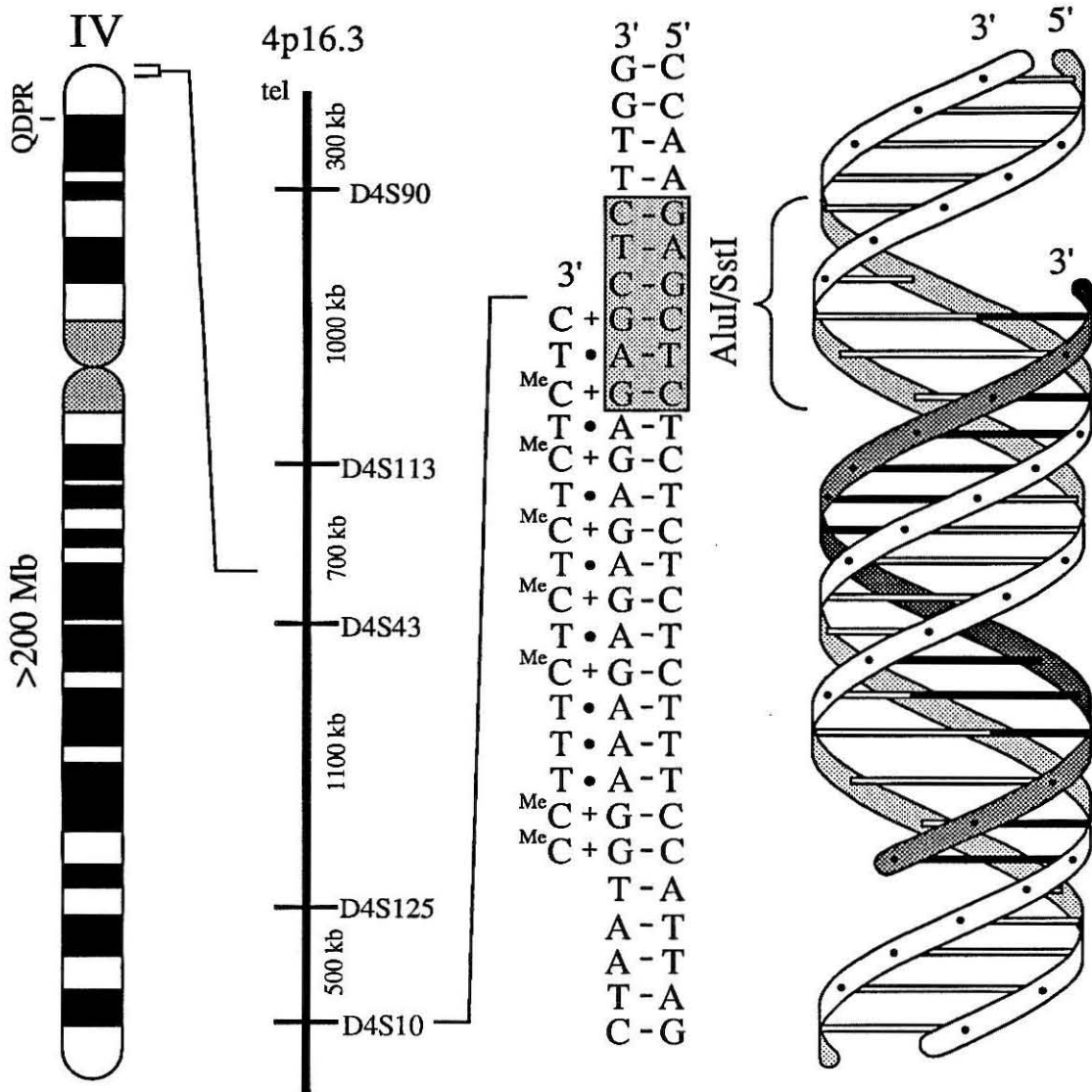


Fig. 4.16. (Left) Schematic diagram of human chromosome 4 showing the cytogenetic location of the QDPR locus at 4p15.3. (Left-Center) A physical map of 4p16.3 to 4pter showing the location and distance between the loci described. On the basis physical mapping, the size estimate from D4S10 to the telomere is approximately 4 Mb. The distance indicated between the loci are approximations from the physical map. (Right-Center) The sequence and relative orientation of the triple helix/AluI-SstI target site. The site is shown being recognized by oligonucleotide Alu-16. (Right) Schematic diagram of the triple helix complex. The pyrimidine oligonucleotide is bound in the major groove, parallel to the purine strand of the DNA duplex and covers the purine half of the AluI methylation site.

of the AluI site was determined by designing a primer from the sequence downstream of the homopyrimidine site and sequencing back in the 5' direction. This demonstrated that the homopurine sequence overlapped a complete AluI site as expected, but it also showed that the AluI site was contained within an SstI site, a restriction enzyme sensitive to AluI methylation (78). This meant that the target site could be cut by site specific inhibition of AluI methylase followed by digestion with the 6-bp restriction enzyme SstI (Fig. 4.16). Interestingly, the target site mapped to within 250 bp of HindIII polymorphism #1, one of the RFLP's originally used to map the chromosomal location of Huntington's disease (24, 71).

Sequence Analysis of the Target Site. The sequence of the 16 bases following the AluI site (those recognized by the degenerate oligonucleotide pool) contained seven G's, eight A's, and one T in addition to the 5'-AG-3' nondegenerate segment within the AluI site. This sequence would be particularly difficult to detect because both the degenerate nucleotide ratio and the oligonucleotide affinity were biased against high G content target sites. At 10 μ M total oligonucleotide, the concentration capable of binding the sequence with a single mismatch at the TA bp was less than 0.3 nM. If one additional random base mismatch was tolerated, for a total of two, the binding concentration would be increased to approximately 4 nM. The cosmid DNA concentration was 0.4 nM giving an oligonucleotide to target ratio of between 1:1 and 10:1 which explains the low cleavage efficiency. This demonstrates that the efficiency of cleavage using degenerate oligonucleotides depends upon both the G content and the number of pyrimidine bases in the candidate sequence.

Conclusion.

The generalizability of triple helix formation is sufficient to eliminate the need for artificial target site insertion into a genetic marker. However, cleavage of endogenous target sites is only possible if sites of the appropriate sequence composition can be identified. While such sites are readily recognized if complete sequence information about a genetic marker is available, little sequence information is known for most marker clones. Although complete sequence determination is a possible strategy, this chapter described three alternatives for a more rapid identification of triple helix target sites. Using the test case of an unsequenced cosmid, 8C10I5, that maps centromeric to the HD mutation, a potential triple helix target site was detected by PCR analysis with partially degenerate oligonucleotide primers, and the sequence identified using enzymatic cleavage by partially degenerate oligonucleotide directed triple helix formation followed by subcloning and sequencing. The target site maps to a region on human chromosome 4 that, if cut, would liberate a single 3-4 Mb fragment containing the entire HD candidate region. The techniques for target site identification are generalizable to any cloned genetic marker.

Part II

Site Specific Cleavage of Human Chromosome 4

The identification of a target site within D4S10 facilitated the design of an oligonucleotide to mediate site specific cleavage of human chromosome 4 at an endogenous target site. The enzyme system to be used for cleavage had not previously been tested on genomic DNA, so it was necessary to optimize the cleavage conditions on plasmid and yeast DNA. Additionally, the use of AluI methylase appeared initially to be problematic due to the higher than statistical frequency of AluI sites in the highly repetitive interspersed repeats found throughout the human genome (79, 80). It was necessary to confirm the target site on human chromosome 4 because the site contained a small repetitive element that might be polymorphic in the human population. Each of these experimental barriers had to be overcome to site specifically cut a human chromosome. Cleavage reactions on human chromosome 4 were described in the report: S. A. Strobel, L. A. Doucette-Stamm, L. Riba, D. E. Housman, and P. B. Dervan; Site-Specific Cleavage of Human Chromosome 4 Mediated by Triple Helix Formation, *Science* 254, 1639-42 (1991).

Confirmation of Target Sequence in Human Chromosome 4. Because the site identified included a simple sequence repeat, (AG)₆, there was concern about the possibility of DNA sequence polymorphism within the targeted region (81). Using PCR with primers flanking the SstI site (SSG8-1 and SSG8-4) the target region from ten unrelated individuals were analyzed including an HD patient, and a somatic cell hybrid (HD113.2B) containing only human chromosome 4 in a mouse background (49). All eleven samples yielded a 213 bp fragment demonstrating there was no variation in the AG repeat length. All samples were subject to digestion with SstI. Sequencing of

several of these PCR products generated the identical sequence to that derived from the 8C10I5 cosmid. This suggests that neither the SstI site nor the (AG)₆ repeat sequence were polymorphic, and DNA from any of these cell lines could serve as substrate for the cleavage reaction.

Target Site Cleavage in Plasmid and Yeast DNA. The ability of an oligonucleotide to effectively block the AluI target site from methylation was tested in plasmid and yeast genomic DNA to optimize methylation conditions and cleavage efficiencies. Plasmid pUCLEU2E was constructed by inserting an oligonucleotide duplex containing the target sequence into the unique XhoI site of pUCLEU2 (3). Quantitative cleavage (>98%) of pUCLEU2E was observed up to pH 7.0 using a 16 base oligonucleotide, Alu-16, containing MeC and T nucleotides (Fig. 4.17, lanes 1-3). A 19 base oligonucleotide (Alu-19G) designed to recognize three additional base pairs at the 5' end of the purine run including a TA bp (82), produced identical cleavage efficiencies under the conditions tested (lanes 4-6). Both oligonucleotides showed reduced cleavage efficiency at higher pH's (lanes 3 & 6). No cleavage was observed at either of the other two SstI sites on this plasmid (compare lane 7), nor was cleavage observed in the absence of the oligonucleotide (lane 10).

To test for cleavage of genomic DNA, the target site was homologously recombined into the *Saccharomyces cerevisiae* genome which contains 14 Mb of chromosomal DNA (3). One recombinant was selected and fully characterized. Unfortunately, this clone contained the target site as an insert into either chromosome VII or XV instead of the desired chromosome III. The chromosome did contain an insertion of the *LEU2* gene near the target site, so it was possible to assay for cleavage by hybridization with *LEU2* (83). As a function of pH, cleavage efficiencies of 80-90% were observed at the target site within megabase DNA (Fig. 4.18). Ethidium bromide staining of

total genomic DNA detected no secondary sites with either oligonucleotide at pH 7.0.

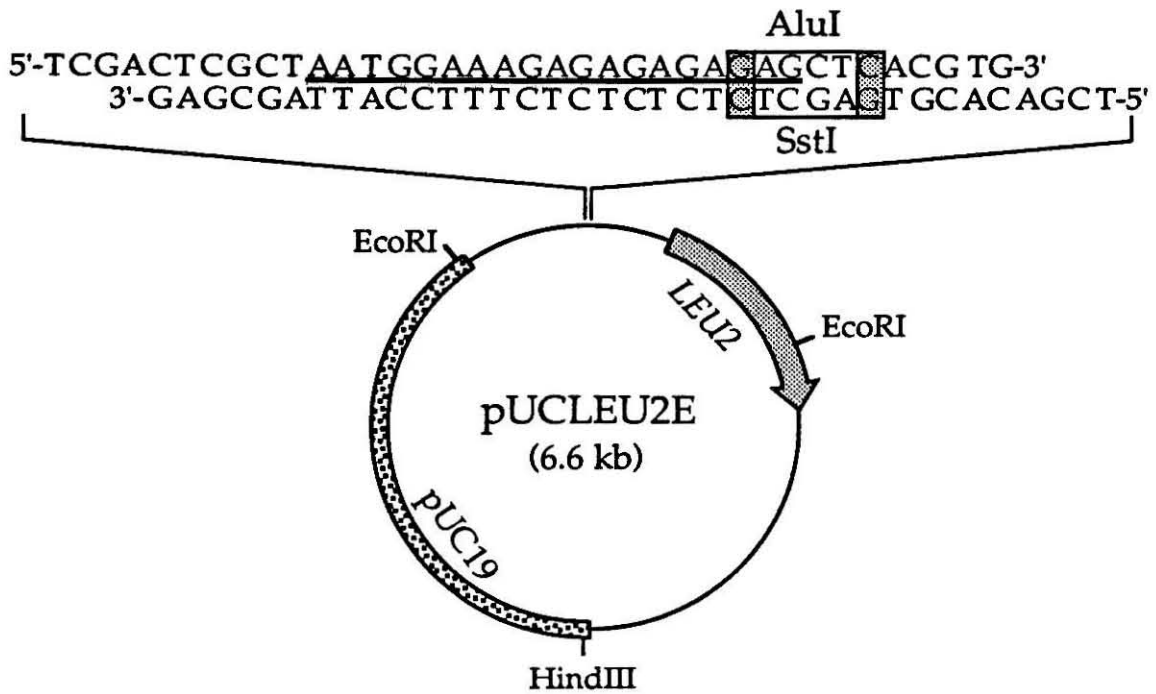
Fig. 4.17. Cleavage of human chromosome 4 target site in pUCLEU2E plasmid DNA. **A.** Sequence of oligonucleotides specific for target site. **B.** Map of plasmid pUCLEU2E showing the duplex ligated into the unique XhoI site. The underlined region in the duplex is the sequence recognized by oligonucleotide Alu-19G. Alu-16 recognizes a sequence three base pairs shorter at the 5' end. Both oligonucleotides partially overlap an AluI site. **C.** Enzymatic cleavage of pUCLEU2E DNA mediated by oligonucleotides Alu-16 and Alu-19G. Plasmid DNA (450 ng) was linearized with HindIII and incubated with 1.0 μ M oligonucleotide Alu-16 or Alu-19G in 1 mM spermine, 50 mM NaCl, 50 mM tris-HCl, pH 6.6-7.4, 10 mM EDTA, and 1 mM 2-mercaptoethanol for 12-14 hours at 22° C. S-adenosyl methionine (160 μ M), BSA (100 μ g/ml) and AluI methylase (5 units) were added to the reaction (final volume 20 μ l) and incubated at 22° for 1.5 hours. The DNA was phenol extracted, ethanol precipitated, 70% ethanol washed, dried, and resuspended in SstI restriction enzyme buffer containing 50 mM NaCl, 50 mM tris-HCl, 10 mM MgCl₂, pH 9.5 (25° C) and 10 units SstI restriction enzyme. Digested the DNA to completion at 37° C for one hour. Resolved products on a 0.8% agarose gel in 1.0x TBE containing ethidium bromide. Visualized products by UV excitation. Intact linearized plasmid DNA is 6.6 kb and the desired cleavage products are each 3.3 kb in size.

A.

Alu-16 5- MeC^{MeC} TTT MeC T^{MeC} T^{MeC} T^{MeC} T^{MeC} T^{MeC} T C -3'

Alu-19G 5- T T G MeC^{MeC} TTT MeC T^{MeC} T^{MeC} T^{MeC} T^{MeC} T^{MeC} T C -3'

B.



C.

	pH	6.6	7.0	7.4	6.6	7.0	7.4	7.0	7.0	7.0	7.0
Alu-16		+	+	+	-	-	-	+	-	-	-
Alu-19G		-	-	-	+	+	+	-	-	-	-
AluI Methylase		+	+	+	+	+	+	-	-	+	+
SstI Endonuclease		+	+	+	+	+	+	+	-	-	+
		1	2	3	4	5	6	7	8	9	10

6.6 kb →
3.3 kb →

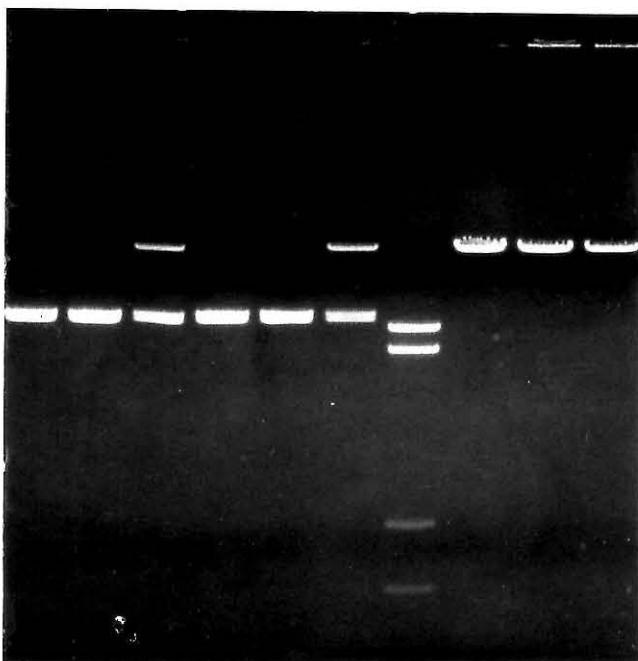
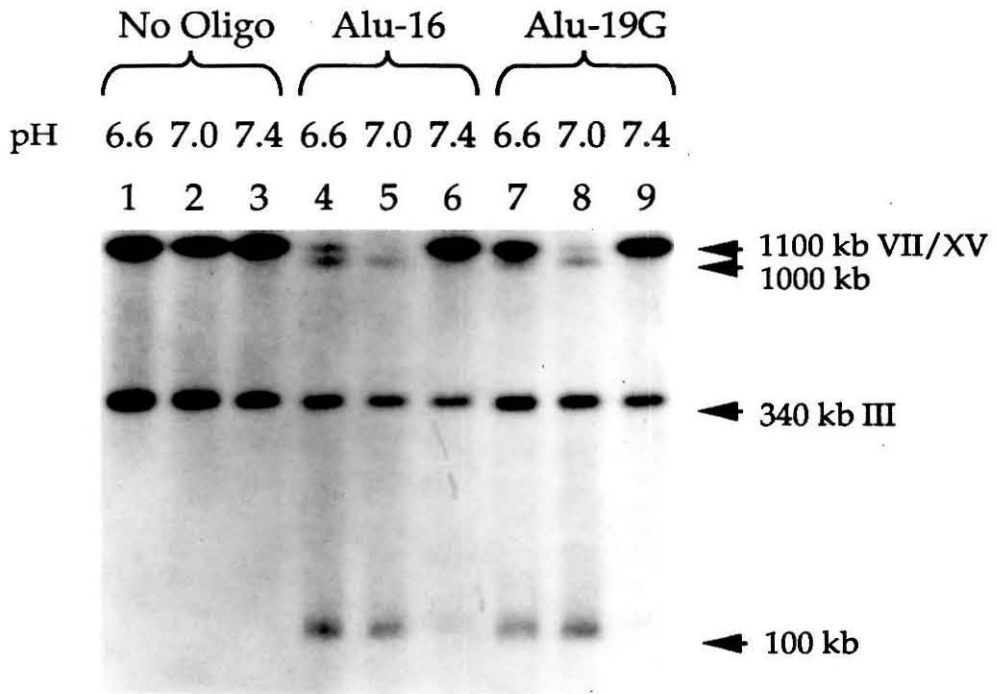


Fig. 4.18 Cleavage of yeast genomic DNA (SEY6210E) containing the human chromosome 4 SstI/AluI target site on chromosome VII or XV. Yeast genomic DNA in LMP agarose blocks ($\approx 80 \mu\text{l}$) was washed twice in 1.0 ml of 50 mM NaCl, 50 mM tris-HCl, pH 6.6-7.4, 10 mM EDTA, 1 mM spermine, and 1 mM 2-mercaptoethanol for 10 min., decanted, and overlaid with 120 μl of the same buffer. Added 2.0 μl of 100 μM oligonucleotide Alu-16 or Alu-19G and incubated 24 hours at 22° C. Added BSA (100 $\mu\text{g}/\text{mL}$) SAM (160 μM) and AluI methylase (32 U) (final concentrations) and incubated an additional 8 hours at 22° C. Inactivated methylase and disrupted triplex by adding 1.0 mL of 1% lauryl sarcosyl, 100 mM tris-HCl pH 9.5, 10 mM EDTA and incubating at 55° C for 30 min. The oligonucleotide and detergent were removed with 4x30 min. washes (1.0 ml) with 10 mM tris-HCl pH 9.5, 10 mM EDTA. The agarose plug was washed twice for 30 min. with 1.0 ml restriction enzyme buffer containing 50 mM NaCl, 50 mM tris-HCl, 10 mM MgCl_2 , pH 8.0, decanted, and overlaid with 120 μl of the same buffer containing 20 units SstI restriction enzyme. Digested at 37° C for 2 hours and resolved products of 1% agarose pulsed field gel containing 0.5x TBE, 120° switch angle, 6 V/cm, 14° C, at switch times ramped from 15-45 sec. for 15 hours followed by a ramp from 60-90 sec. for 5 hours. DNA was transferred to Nytran membrane and hybridized with the KpnI/EcoRI fragment from the *LEU2* gene as described in Chapter 3.



cyglycerol by such a mechanism could implicate protein kinase C as a possible transducing pathway. Or, mobilization of intracellular Ca^{2+} by inositol triphosphate could point to Ca^{2+} -calmodulin-dependent protein (CaM) kinases as the transducing factors that couple dopamine stimulation of D_1 receptors to the activation of steroid receptors. CaM kinases phosphorylate and activate the cAMP response element binding-protein (33).

Demonstration of "cross-talk" between membrane-associated receptors and intracellular steroid hormone receptors may be of biomedical significance. Dual activation of receptors might occur in situ in brain cells. It may also be possible to activate mutant forms of steroid receptors that exist in diseases (34). Finally, the finding that dopamine has direct access to the genome via this family of transcription factors may aid in understanding learning and memory processes.

REFERENCES AND NOTES

- K. R. Yamamoto, *Annu. Rev. Genet.* **19**, 209 (1985); R. M. Evans, *Science* **240**, 889 (1988); S. Green and P. Chambon, *Trends Genet.* **4**, 309 (1988); M. Beato, *Cell* **56**, 335 (1989); B. W. O'Malley, *Mol. Endocrinol.* **4**, 363 (1990).
- J. Gorski, D. Toft, G. Shyamala, D. Smith, A. Norides, *Recent Prog. Horm. Res.* **24**, 45 (1968); E. V. Jensen et al., *Proc. Natl. Acad. Sci. U.S.A.* **59**, 632 (1968).
- B. W. O'Malley, W. L. McGuire, P. O. Kohler, S. Korenman, *Recent Prog. Horm. Res.* **25**, 105 (1969).
- T. Steimer and J. B. Hutchison, *Nature* **292**, 345 (1981); B. Parsons, N. J. MacLusky, L. Krey, D. W. Pfaff, B. S. McEwen, *Endocrinology* **107**, 774 (1980); E. R. deKloet, T. A. M. Voorhuis, J. Elands, *Eur. J. Pharmacol.* **118**, 185 (1985).
- J. D. Blaustein, H. I. Ryer, H. H. Feder, *Neuroendocrinology* **31**, 403 (1980); L. Bogic, J. L. Gerlach, B. S. McEwen, *Endocrinology* **122**, 2735 (1988).
- S. A. Sholl and S. M. Pomerantz, *Endocrinology* **119**, 1625 (1986).
- M. Schumacher, *Trends Neurosci.* **13**, 359 (1990).
- C. Beyer, E. Canchola, K. Larsson, *Physiol. Behav.* **26**, 249 (1981).
- L. A. Denner, N. L. Weigel, B. L. Maxwell, W. T. Schrader, B. W. O'Malley, *Science* **250**, 1740 (1990).
- C. Glineur, M. Zenke, H. Beuq, J. Ghysdael, *Genes Dev.* **4**, 1663 (1990).
- L.-H. Wang et al., *Nature* **340**, 163 (1989).
- R. F. Power, J. P. Lydon, O. M. Conneely, B. W. O'Malley, *Science* **252**, 1546 (1991).
- O. M. Conneely, D. M. Kettelberger, M.-J. Tsai, W. T. Schrader, B. W. O'Malley, *J. Biol. Chem.* **264**, 14062 (1989).
- H. N. Jantzen et al., *Cell* **49**, 29 (1987).
- N. L. Weigel and B. W. O'Malley, unpublished data.
- R. F. Power, O. M. Conneely, B. W. O'Malley, unpublished data.
- The human ER expression construct pSVMt-wER contains the wild-type human ER cDNA inserted downstream of a metallothionein promoter. The ERE1bCAT reporter plasmid contains a fragment of the vitellogenin gene promoter (-331 to -87) upstream of an ElbTATA box linked to a CAT gene.
- V. Giguere, S. M. Hollenberg, M. G. Rosenfeld, R. M. Evans, *Cell* **46**, 645 (1986).
- J. L. Arriza et al., *Science* **237**, 268 (1987).
- M. S. Bradshaw, M.-J. Tsai, B. W. O'Malley, *Mol. Endocrinol.* **2**, 1286 (1988).
- S. K. Nordeen, B. Kuhnel, J. Lawler-Heavner, D. A. Barber, D. P. Edwards, *ibid.* **3**, 1270 (1989).
- J. Kebabian and D. Calne, *Nature* **277**, 93 (1979); J. R. Bunzow et al., *ibid.* **336**, 783 (1988); A. Dearny et al., *ibid.* **347**, 72 (1990); P. Sokoloff, B. Giros, M.-P. Martres, M.-L. Bouthenet, J.-C. Schwartz, *ibid.*, p. 146; H. H. M. Van Tol et al., *ibid.* **350**, 610 (1991); R. K. Sunahara et al., *ibid.*, p. 614.
- R. F. Power, O. M. Conneely, B. W. O'Malley, data not shown. The results represent mean values obtained in two independent experiments.
- A. D. W. Dobson et al., *J. Biol. Chem.* **264**, 4207 (1989).
- L. A. Denner, W. T. Schrader, B. W. O'Malley, N. L. Weigel, *J. Biol. Chem.* **265**, 16548 (1990).
- L. A. Denner, N. L. Weigel, B. W. O'Malley, unpublished data.
- A. Guichon-Mantel et al., *Cell* **57**, 1147 (1989); J. A. Simental, M. Sar, M. V. Lane, F. S. French, E. M. Wilson, *J. Biol. Chem.* **266**, 510 (1991). The difference in results may reflect that, in our studies, we replaced charcoal-stripped serum with Nutridoma in the growth media 24 hours prior to transfection. This would eliminate the carryover of progesterone, dopamine, and serum-stimulable phosphorylation factors, among others, which may be present in trace amounts even in charcoal-stripped serum.
- C. E. Reker, M. C. LaPointe, B. Kovacic-Milivojevic, W. J. H. Chiu, W. V. Vedeckis, *J. Steroid Biochem.* **26**, 653 (1987).
- P. L. Mobley and F. Sulser, *Nature* **286**, 608 (1980); P. L. Mobley, D. H. Manier, F. Sulser, *J. Pharmacol. Exp. Ther.* **226**, 71 (1983); F. Sulser, *J. Clin. Psychiatr.* **10**, 13 (1986).
- A. Aperia et al., *Am. J. Hypertens.* **3**, 11S (1990).
- K. R. Williams, H. C. Hemmings, M. B. LoPresti, W. H. Konigsberg, P. Greengard, *J. Biol. Chem.* **261**, 1890 (1986); P. Stralifors, H. C. Hemmings, P. Greengard, *Eur. J. Biochem.* **180**, 143 (1989).
- C. C. Felder, M. Belcher, P. A. Jose, *J. Biol. Chem.* **264**, 8739 (1989).
- M. Sheng, M. A. Thompson, M. E. Greenberg, *Science* **252**, 1427 (1991).
- M. R. Hughes et al., *ibid.* **242**, 1702 (1988).
- C. M. Gorman, L. F. Moffatt, B. M. Mavard, *Mol. Cell. Biol.* **2**, 1044 (1982).
- The experiments shown were authenticated with several batches of CV₁ cells from different sources. We observed with some batches of CV₁ cells that the dopamine response began to disappear after several successive passages for unknown reasons. In these cases, we could restore the response by changing to cells that had not been passaged as frequently.
- M. J. Toro, E. Montaya, L. Birnbaumer, *Mol. Endocrinol.* **1**, 669 (1987).
- G. G. Wong et al., *Science* **228**, 810 (1985).
- The LD₁ mutant was constructed by use of oligonucleotide site-directed mutagenesis as described (13, 24). We prepared the cDNA template for mutagenesis by subcloning a 5'-Pst I-Sst I fragment of the receptor cDNA into M13 mp18. We subcloned mutants by synthesizing a second strand on the mutated M13 template with Klenow DNA polymerase, cutting with Sst I and Pst I, and ligating to an Sst I-Pst I digested plasmid (SP₄) (40) that contained the wild-type PR_A sequence. We confirmed the mutated sequence by dideoxy sequencing [F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977)].
- O. M. Conneely, B. L. Maxwell, D. O. Toft, W. T. Schrader, B. W. O'Malley, *Biochem. Biophys. Res. Commun.* **149**, 493 (1987).
- W. C. Okulicz, A. M. Savasta, L. M. Hoberg, C. Longcope, *Endocrinology* **125**, 930 (1989).
- W. P. Sullivan, T. G. Beito, J. Proper, C. J. Krco, D. O. Toft, *ibid.* **119**, 1549 (1986).
- We thank P. Kushner and G. L. Greene for pSVMt-wER, V. Allgood and J. Cidlowski for ERE1bCAT, R. M. Evans for pRShGR α , J. L. Arriza for pRShMR, and S. Nordeen for pAHCAT. We thank D. O. Toft for antibody PR22, and D. Gallup, K. Jackson, and D. Scott for technical assistance. We also thank L. Gamble and D. Scarff for help in the preparation of the manuscript and figures, respectively.

12 June 1991; accepted 9 September 1991

Site-Specific Cleavage of Human Chromosome 4 Mediated by Triple-Helix Formation

SCOTT A. STROBEL, LYNN A. DOUCETTE-STAMM, LAURA RIBA, DAVID E. HOUSMAN, PETER B. DERVAN*

Direct physical isolation of specific DNA segments from the human genome is a necessary goal in human genetics. For testing whether triple-helix mediated enzymatic cleavage can liberate a specific segment of a human chromosome, the tip of human chromosome 4, which contains the entire candidate region for the Huntington's disease gene, was chosen as a target. A 16-base pyrimidine oligodeoxyribonucleotide was able to locate a 16-base pair purine target site within more than 10 gigabase pairs of genomic DNA and mediate the exact enzymatic cleavage at that site in more than 80 percent yield. The recognition motif is sufficiently generalizable that most cosmids should contain a sequence targetable by triple-helix formation. This method may facilitate the orchestrated dissection of human chromosomes from normal and affected individuals into megabase sized fragments and facilitate the isolation of candidate gene loci.

PYRIMIDINE OLIGODEOXYRIBONUCLEOTIDES bind in the major groove of DNA parallel to the purine Watson-

Crick strand through formation of specific Hoogsteen hydrogen bonds to the purine Watson-Crick base (1-6). Specificity is derived from thymine (T) recognition of adenine·thymine (AT) base pairs (TAT triplet); and N³-protonated cytosine (C⁺) recognition of guanine·cytosine (GC) base pairs (C + GC triplet) (1-8). The generalizability of triple-helix formation has been extended beyond purine tracts to mixed

S. A. Strobel and P. B. Dervan, Arnold and Mabel Beckman Laboratories of Chemical Synthesis, California Institute of Technology, Pasadena, CA 91125.
L. A. Doucette-Stamm, L. Riba, D. E. Housman, Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139.

*To whom correspondence should be addressed.

sequences containing all four base pairs by the identification of novel base triplets (9), alternate strand triple-helix formation (10), and other triple-helix motifs (11). By combining oligonucleotide-directed recognition with enzymatic cleavage, near quantitative cleavage at a single target site within megabase (Mb) DNA has been achieved (12). In this technique, specific triple-helix formation protects a single overlapping methylation site from modification (12, 13). After triple-helix disruption, restriction enzyme digestion produced cleavage in 95% yield at a single site within the *Saccharomyces cerevisiae* genome (14 Mb) (12).

In a formal sense, a binding site size of at least 16 bp afforded by the triple helix motif is sufficient to target millions of

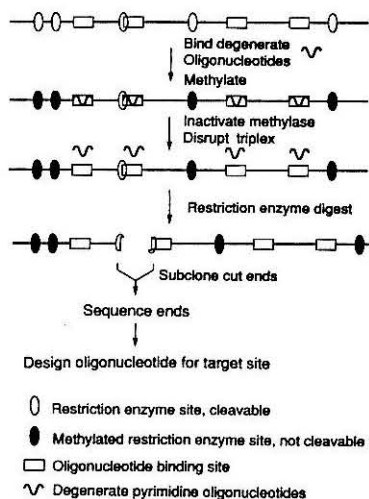


Fig. 1. General scheme for identification of endogenous cosmid sequences targetable by triple-helix mediated enzymatic cleavage with the use of a pool of partially degenerate oligonucleotides. Linearized cosmid DNA was equilibrated with a degenerate pyrimidine oligonucleotide containing a nondegenerate end specific for a methylase site. The Alu I oligonucleotide used to identify this target site had the sequence, 5'-Y₁₆TC-3', where Y represents a nucleotide ratio of 60% ^BU and 40% ^MC. The DNA was methylated to completion, making all methylation sites not bound by the oligonucleotide mixture resistant to subsequent restriction enzyme digestion. The methylase was removed by phenol extraction, the DNA was precipitated with ethanol, and the triple helix was disrupted by resuspension in a restriction enzyme buffer at high pH. The appearance of a cleavage product upon restriction enzyme digestion demonstrated that a site bound by a fraction of the oligonucleotide mixture overlaps a methylation site. Subcloning of the cut ends and sequencing revealed the site identity. With this sequence information and the rules for oligonucleotide directed recognition of duplex DNA (1-6, 9, 10), an oligonucleotide specific for the target site was designed.

different sites, each occurring only once in gigabase DNA (14). To test whether triple helix-mediated enzymatic cleavage is capable of site-specific cleavage at an endogenous sequence on a human chromosome, we targeted the tip of the short arm of chromosome 4 (>200 Mb), which contains the Huntington's disease (HD) gene. The HD gene is located within the cytogenetic band 4p16.3, and a large number of cloned DNA sequences have been genetically and physically mapped to the region (15-16). Genetic evidence supports a location of the HD gene within the approximately 4 Mb of DNA between the genetic marker D4S10 (G8) (15) and the telomere of chromosome 4p (17, 18). Thus, a single cleavage at D4S10 would liberate a large, resolvable fragment that contains the entire HD candidate region.

The triple helix-recognition motif is sufficiently general that each cosmid should contain at least one endogenous site suitable for enzymatic cleavage (19). For cosmid clones that have not been sequenced, a strategy was developed whereby a pool of partially degenerate pyrimidine oligonucleotides was used to identify target sites (Fig. 1). Pyrimidine oligonucleotides were designed with two domains; a degenerate 16-nucleotide (nt) segment capable of binding thousands of purine-rich sequences, and a short specific 2- to 3-nt domain to overlap a methylase-restriction enzyme binding site. The degenerate oligonucleotides were synthesized with 5-methylcytosine (^MC) and 5-bromouracil (^BU)

(4). These analogs were chosen because ^MC substituted oligonucleotides have higher binding affinities and ^BU substitutions result in slightly reduced binding specificity (4, 12). Sequences containing high G content or contiguous G tracts are not efficiently targeted by the pyrimidine triplex motif near neutral pH. Because oligonucleotides in the population specific for such sequences would be less likely to bind, a nucleotide ratio of 60% ^BU and 40% ^MC at each degenerate position was used to minimize the oligonucleotide concentration specific for G-rich sequences. This resulted in a pool of 65,536 different oligodeoxyribonucleotide sequences unequally represented in the population (20).

In an effort to identify potential target sites within D4S10, cosmid clone 8C1015, which contains most of the D4S10 locus (15), was screened with five partially degenerate pyrimidine oligonucleotides in combination with five methylase-restriction enzyme sets (Fig. 1). While no detectable cleavage was observed with Eco RI, Msp I, Hae III, or Taq I, a degenerate oligonucleotide specific for Alu I blocked a single Alu I methylation site in 8C1015. DNA sequencing immediately adjacent to the cleavage site identified the target as a 16-bp purine sequence that overlapped the purine half sites of Alu I methylase (5'-AGCT-3') and Sst I (5'-GAGCTC-3'), a restriction enzyme sensitive to Alu I methylation (Fig. 2). The ability of a target-specific oligonucleotide to effectively bind and fully protect this site was optimized on

Fig. 2. (Left) Schematic diagram of human chromosome 4 showing the cytogenetic location of the QDPR locus at 4p15.3. (Left-Center) A physical map of 4p16.3 to 4p16ter showing the location and distance between the loci described. On the basis of pulsed field gel mapping, the size estimate from D4S10 to the telomere is approximately 4 Mb (17). The distances indicated between the loci are approximations from the physical map (17). (Right-Center) The sequence and relative orientation of the Alu I-Sst I target site identified by the degenerate pyrimidine oligonucleotide search. The site is recognized by a 16-base ^MCCT oligonucleotide with ^MC + GC and TAT base triplets (Alu-16). (Right) Schematic diagram of the triple-helix complex. The pyrimidine oligonucleotide is bound in the major groove, parallel to the purine strand of the DNA duplex and covers the purine half of the Alu I methylation site.

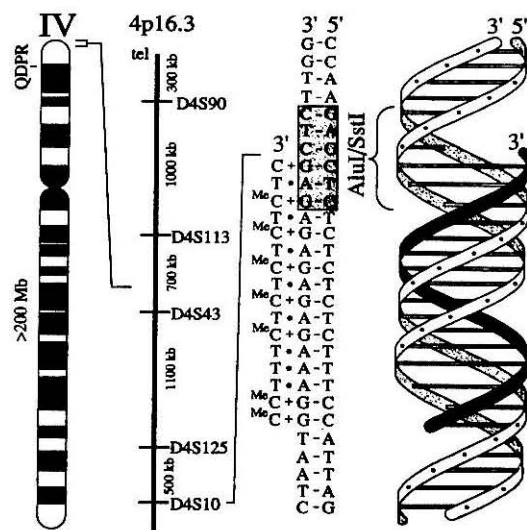
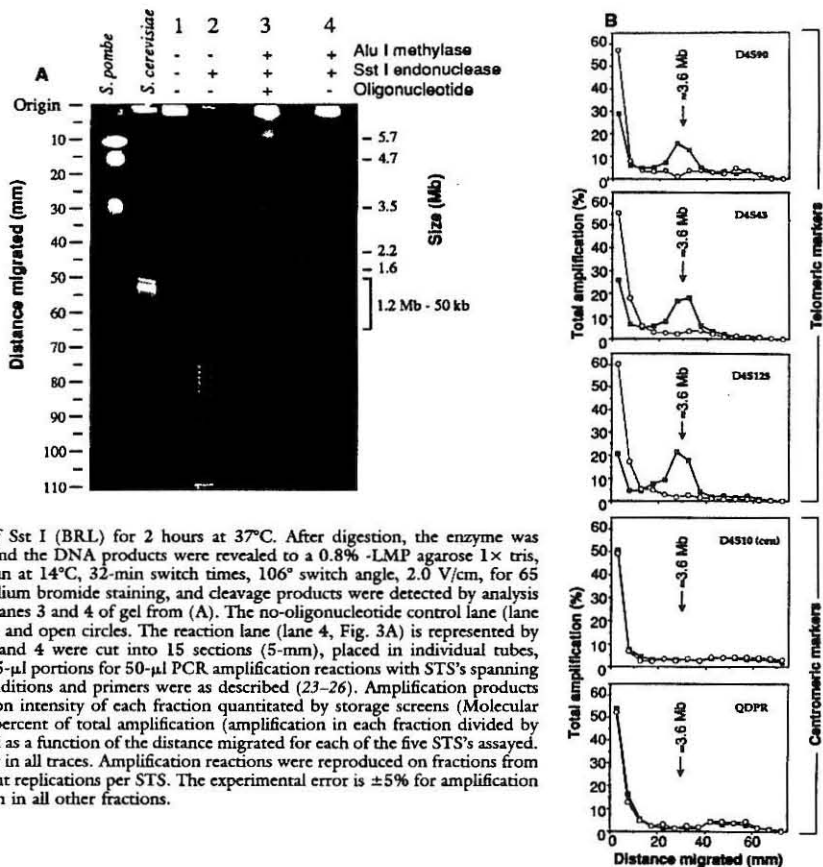


Fig. 3. (A) Triple-helix mediated enzymatic cleavage of human chromosome 4 with reagents indicated above figure. The DNA was prepared from somatic cell hybrid HD113.2B in low melting point (LMP) agarose at ≈ 100 $\mu\text{g/ml}$ by standard procedures (27). Approximately 80 μl of agarose embedded genomic DNA (≈ 8 μg) was equilibrated in triple-helix-Alu I methylase buffer (50 mM NaCl, 1 mM spermine-4HCl, 50 mM tris-HCl, 10 mM EDTA, 1 mM β -mercaptoethanol, pH 7.0) and incubated for 24 hours at room temperature with 5 μM Alu-16 target-specific oligonucleotide. S-Adenosylmethionine (SAM) (160 μM) and bovine serum albumin (BSA) (100 $\mu\text{g/ml}$) were added, and the DNA was methylated to completion with 25 units of Alu I methylase (New England Biolabs) at room temperature for 8 hours. Methylase inactivation and triplex disruption were as described (3). The DNA was reequilibrated in Sst I restriction enzyme buffer (50 mM NaCl, 50 mM tris-HCl, pH 7.9, and 10 mM MgCl_2) and cut to completion with 20 units of Sst I (BRL) for 2 hours at 37°C. After digestion, the enzyme was inactivated by heat for 10 min at 55°C and the DNA products were revealed to a 0.8% -LMP agarose 1 \times tris, acetate, EDTA (TAE) pulsed-field gel run at 14°C, 32-min switch times, 106° switch angle, 2.0 V/cm, for 65 hours. Reactions were visualized by ethidium bromide staining, and cleavage products were detected by analysis with the PCR (23). **(B)** PCR analysis of lanes 3 and 4 of gel from (A). The no-oligonucleotide control lane (lane 3, Fig. 3A) is represented with gray lines and open circles. The reaction lane (lane 4, Fig. 3A) is represented by black lines with filled squares. Lanes 3 and 4 were cut into 15 sections (5-mm), placed in individual tubes, denatured at 96°C, vortexed, and used in 5- μl portions for 50- μl PCR amplification reactions with STS's spanning the 4p16.3 region. The amplification conditions and primers were as described (23-26). Amplification products were blot hybridized and the amplification intensity of each fraction quantitated by storage screens (Molecular Dynamics 400S PhosphorImager). The percent of total amplification (amplification in each fraction divided by total amplification in the lane) was plotted as a function of the distance migrated for each of the five STS's assayed. The lane origin is the first point to the left in all traces. Amplification reactions were reproduced on fractions from four different gels, resulting in two to eight replications per STS. The experimental error is $\pm 5\%$ for amplification at each origin, and $\pm 2\%$ for amplification in all other fractions.



plasmid (6.6 kb) and yeast genomic DNA (14 Mb), resulting in more than 98% and 90% cleavage yields at the target site, respectively. No secondary cleavage was detected at other Sst I sites in either substrate. Thus, after triple-helix formation and exhaustive Alu I methylation, the megabase genomic DNA could be cut with Sst I to reveal only the site bound by the oligonucleotide (Fig. 2).

Under optimized conditions, triple-helix mediated enzymatic cleavage was used to target human chromosome 4 in DNA prepared from HD113.2B, a somatic cell hybrid containing only human chromosome 4 in a mouse background (21) (Fig. 3A). Analysis of the reaction products within the range of 1 to 6 Mb by pulsed-field gel electrophoresis (22), revealed that most of the DNA did not migrate from the origin and remained intact (lanes 1, 3, and 4). When methylation was omitted and the DNA was digested with Sst I, no DNA remained at the origin, a result consistent

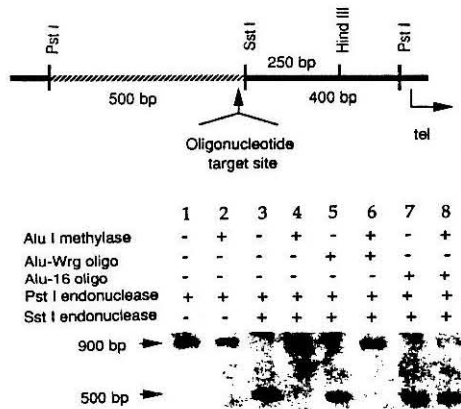
with complete digestion (lane 2). Reactions in the presence (lane 3) or absence (lane 4) of oligonucleotide appeared identical by ethidium bromide staining. Because the maximum yield of the telomeric segment of chromosome 4 was 2 ng (from ≈ 8 μg of total DNA), any cleavage product would be undetectable with ethidium bromide staining, thus requiring a more sensitive detection system.

Southern blot hybridization revealed a discrete product approximately 3.5 Mb in size. To improve the sensitivity of detection, a PCR (polymerase chain reaction) based assay to detect small quantities of specific megabase DNA segments was applied to the gel products shown in Fig. 3A (Fig. 3B). The full length of lanes 3 (with oligonucleotide) and 4 (without oligonucleotide) were cut into 5-mm sections of the gel and placed into individual tubes. Each fraction was analyzed for genomic DNA content by the PCR with the use of primer sets specific

for sequence tagged sites (STS) known to be either telomeric or centromeric to the target site. Centromeric primers would be expected to only amplify the gel fraction containing the origin because the main portion of the chromosome would be too large to migrate into the gel. Telomeric markers, however, would amplify both the origin and any fractions containing the telomeric product expected from restriction mapping to be about 4 Mb in size (17, 18).

Amplification of D4S90 (23), the most telomeric 4p marker (Fig. 2) (16, 17), was detected in the origin of both lanes (Fig. 3B). The amplification signal was weaker in the origin of lane 3, consistent with chromosomal cleavage. Significant amplification within lane 3 was also observed in fractions 6 and 7, while the control lane containing no oligonucleotide remained baseline in all other fractions (lane 4). Fractions 6 and 7 correspond to a migration distance of 25 to 35 mm and correlate

Fig. 4. The yield of triple helix-mediated enzymatic cleavage of human chromosome 4. (**Upper**) Local restriction map showing the location of the Alu I-Sst I target site within the 900-bp Pst I fragment. The target site is located 250 bp centromeric to Hind III polymorphic site no. 1 (15, 25). (**Lower**) Reaction of total genomic HD113.2B DNA as indicated above figure and described in Fig. 5. Pst I (80 units) (New England Biolabs) was included in the 2-hour Sst I digest to generate the desired fragments. Products were resolved by standard agarose gel electrophoresis (0.8% agarose, 1× TAE), blotted to a Nytran 66 membrane (Schleicher & Schuell) and hybridized with a 500-bp centromeric Pst I-Sst I fragment (striped region) according to the protocol issued with the membrane. The cleavage efficiency was determined from the relative radioactive intensities of the 500- and 900-bp products as determined by PhosphorImager analysis.



to a size estimate of 3.6 ± 0.3 Mb, on the basis of a comparison to the size standard from *Saccharomyces pombe*. A similar pattern of amplification was observed for STS's from D4S43 (16, 23) and D4S125 (24). These loci span the candidate region for the HD gene, and amplification demonstrates that the cleavage product contains all the DNA from D4S10 to the telomere (Fig. 2).

Amplification of STS's centromeric to the cut site showed a different pattern. Approximately equal amplification was detected in the origins of both lanes and only in the first fraction. No peak was detected with a D4S10 STS located 12 kb centromeric to the cut site (25), nor was a peak present with a QDPR STS located in 4p15.3 (26), one cytogenetic band centromeric to D4S10 (Fig. 2). This result demonstrated that these proximal loci are not found in the 3.6-Mb cleavage product, and indicated that contamination was not contributing to the peaks seen with the telomeric STS's.

The cleavage efficiency at the target site was determined directly by DNA (Southern) blotting. The Sst I target site is found near the center of a 900-bp Pst I fragment (Fig. 4, lanes 1 to 3). After Pst I digestion, cleavage at the Sst I site can be readily detected and quantitated by Southern blot hybridization with the 500-bp centromeric fragment (Fig. 4). The ratio of the number of counts at 500 bp to the total at both the 500- and 900-bp sites is the cleavage efficiency. To determine the yield of cleavage at the Sst I site in human chromosome 4, we performed reactions as described in the legend of Fig. 3A except that Pst I was included in the restriction enzyme diges-

tion. No 500-bp cleavage product was observed if oligonucleotide was omitted (lane 4) or if an oligonucleotide with the incorrect sequence was incubated with the DNA (lane 6). The DNA was quantitatively converted to the 500-bp fragment in the absence of Alu I methylase (lanes 3, 5, and 7). In the presence of Alu I methylase and the correct oligonucleotide, the 500-bp diagnostic product was produced in 80 to 90% yield (lane 8). This is only slightly lower than observed at the plasmid level despite the presence of more than 10 billion base pairs of background DNA per target site.

REFERENCES AND NOTES

- H. E. Moser and P. B. Dervan, *Science* **238**, 645 (1987); S. A. Strobel, H. E. Moser, P. B. Dervan, *J. Am. Chem. Soc.* **110**, 7927 (1988).
- D. Praseuth et al., *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1349 (1988); V. I. Lyamichev et al., *Nucleic Acids Res.* **16**, 2165 (1988).
- L. J. Maher III, B. Wold, P. B. Dervan, *Science* **245**, 725 (1989); J. C. Hanvey, M. Shimizu, D. Wells, *Nucleic Acids Res.* **18**, 157 (1990).
- T. J. Povsic and P. B. Dervan, *J. Am. Chem. Soc.* **111**, 3059 (1989).
- S. A. Strobel and P. B. Dervan, *Science* **249**, 73 (1990).
- L. J. Maher, P. B. Dervan, B. Wold, *Biochemistry* **29**, 8820 (1990).
- G. Felsenfeld, D. R. Davies, A. Rich, *J. Am. Chem. Soc.* **79**, 2023 (1957); M. N. Lipsett, *J. Biol. Chem.* **239**, 1256 (1964).
- P. Rajagopal and J. Feigon, *Nature* **239**, 637 (1989); ———, *Biochemistry* **28**, 7859 (1989); C. de los Santos et al., *ibid.*, p. 7282 (1989).
- L. C. Griffin and P. B. Dervan, *Science* **245**, 967 (1989).
- D. A. Horne and P. B. Dervan, *J. Am. Chem. Soc.* **112**, 2435 (1990).
- M. Cooney, G. Czernuszewicz, E. H. Postel, S. J. Flint, M. E. Hogan, *Science* **241**, 456 (1988); P. A. Beal and P. B. Dervan, *Science* **251**, 1360 (1991).
- S. A. Strobel and P. B. Dervan, *Nature* **350**, 172 (1991).
- M. Koob and W. Szybalski, *Science* **250**, 271 (1990).
- P. B. Dervan, in *Nucleic Acids and Molecular Biology*, F. Eckstein and D. M. J. Lilley, Eds. (Springer-Verlag, Heidelberg, 1988), vol. 2, p. 49.
- J. F. Gusella et al., *Nature* **306**, 234 (1983); S. E. Folstein et al., *Science* **229**, 776 (1985); J. E. Landegent et al., *Hum. Genet.* **73**, 354 (1986); J. Gusella et al., *Cold Spring Harbor Symp. Quant. Biol.* **51**, 359 (1986); M. I. Skraastad et al., *Am. J. Hum. Genet.* **44**, 560 (1989).
- S. Youngman et al., *Genomics* **5**, 802 (1989); J. J. Wasmuth et al., *Nature* **332**, 734 (1988); J. E. Richards et al., *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6437 (1988); T. M. Pohl et al., *Nucleic Acids Res.* **16**, 9185 (1988); T. C. Gilliam et al., *Science* **238**, 950 (1987).
- T. C. Gilliam et al., *Cell* **50**, 565 (1987); C. A. Pritchard et al., *Genomics* **4**, 408 (1989); M. Bucan et al., *ibid.* **6**, 1 (1990).
- G. P. Bates et al., *Am. J. Hum. Genet.* **49**, 7 (1991).
- A sequence suitable for triple-helix formation was calculated to occur approximately once every 2000 bp (2). A target sequence that also overlaps a single 4-bp methylase would occur approximately 64 times less frequently, or every 120,000 bp. The use of five commercially available 4-bp methylases compatible with the enzymatic cleavage technique (Alu I, dam, Hae III, Hpa II, and Taq I) increases the frequency of a target site to about one every 24,000 bp, less than the average insert size of a cosmid clone.
- The ^{18}U and ^{13}C nucleotide ratio at each position resulted in an oligonucleotide pool in which 85% of the molecules have an overall base ratio for ^{18}U to ^{13}C of 50% or higher. At 10 μM total oligonucleotide, the $(^{18}\text{U})_{16}$ sequence had the highest concentration at 3 nM, and the $(^{13}\text{C})_{16}$ the lowest at 4 pM. The oligonucleotide composition with the highest total concentration (2 μM total, 0.25 nM in each sequence) were sequences with overall base content $(^{18}\text{U})_{10} (^{13}\text{C})_6$.
- L. A. Doucette-Stamm et al., *Somatic Cell Mol. Genet.*, in press.
- D. C. Schwartz and C. R. Cantor, *Cell* **37**, 67 (1984); G. F. Carle and M. V. Olson, *Nucleic Acids Res.* **12**, 5647 (1984); G. F. Carle, M. Frank, M. V. Olson, *Science* **232**, 65 (1986); S. M. Clark, E. Lai, B. W. Birren, L. Hood, *Science* **241**, 1203 (1988); B. W. Birren, E. Lai, S. M. Clark, L. Hood, M. I. Simon, *Nucleic Acids Res.* **16**, 7563 (1988).
- J. F. Gusella et al., *Genomics*, in press; R. K. Saiki, in *PCR Technology*, H. A. Erlich, Ed. (Stockton Press, New York, 1989), pp. 7-16; H. A. Erlich, D. Gelfand, J. J. Sninsky, *Science* **252**, 1643 (1991).
- Because a nonpolymorphic PCR product was desired for D4S125, primers to the 3' side of the variable terminal repeat were selected from the sequence published by B. Richards, G. T. Horn, J. J. Merrill, K. W. Klinger, *Genomics* **9**, 235 (1991). The primers were 5'-CTCTGGTCTGAGGTGCTGAC-3' and 5'-CGGCAGCTGAGGAGGTGCCT-3'. Reactions were performed at 28 cycles, 94°C, 1 min; 60°C, 1 min; and 72°C, 1 min; and standard buffer conditions as described (23).
- P. M. Stapleton, *Nucleic Acids Res.* **16**, 2735 (1988).
- S. Theune, J. Fung, S. Todd, A. Y. Sakaguchi, S. L. Naylor, *Genomics* **9**, 511 (1991).
- C. L. Smith, S. R. Klco, C. R. Cantor, in *Genome Analysis: A Practical Approach*, K. Davies, Ed. (IRL Press, London, England, 1988), pp. 41-72.
- We thank J. F. Gusella for cosmid 8C1015, and assistance with the physical map location of the target site; the members of the Hereditary Disease Foundation HD Collaborative Research group for STS-PCR primer sequences and helpful discussions; the Howard Hughes Medical Institute for a predoctoral fellowship to S.A.S.; the Joan and William Schreyer Research Fund for postdoctoral fellowship to L.D.S.; the National Institutes of Health (HG00098 and HG00329) and the Hereditary Disease Foundation for grant support.

3 September 1991; accepted 12 November 1991

Complete AluI Methylation of Mammalian Genomic DNA. Extensive effort was made to achieve complete AluI methylation of DNA prepared from the somatic cell hybrid HD113.2B. Initial reactions were performed using the manufacture recommended low salt (0 mM) buffer at pH of 6.6 to 7.4 and 1 mM spermine. This resulted in DNA digested to completion by SstI endonuclease, an observation consistent with very poor methylase activity. Conversely, reactions performed in the absence of spermine showed complete methylation and no sensitivity to SstI. A similar pattern of sensitivity was observed on plasmid DNA. The data was consistent with DNA precipitation by spermine prior to the addition of methylase, thus making it inaccessible to the enzyme (84). Because the chromosomal DNA was suspended in agarose, the high pH washes and reequilibration in restriction enzyme buffer removed the spermine and resuspended the DNA, making it susceptible to digestion by SstI. The precipitation problem was corrected by the addition of 50 mM NaCl to the triple helix/methylation reaction. The sodium ions competed with the spermine for the anionic DNA backbone and kept the chromosomes in solution. The medium salt and low pH reduced the activity of the methylase, thus requiring more units and longer incubation times to achieve complete methylation.

An additional concern was the concentration of the genomic DNA in the agarose blocks. Initial methylation reactions performed on DNA prepared from HD113.2B at 10^8 cells/ml were completely ineffective. The agarose block was too viscous with DNA and cellular debris for the methylase to efficiently diffuse into the matrix. Experiments performed on a range of plugs at different cell concentrations showed a maximum concentration of 7.5×10^6 diploid cells/ml or approximately 80 μ g DNA/ml for efficient methylation.

Concentrations even twice as high showed significantly reduced methylase activity, independent of the number of units of enzyme added.

Analysis of the 3.6 Mb Cleavage Product within the D4S10 Locus. PCR analysis of the gel fractions confirmed that the cleavage product contained DNA from D4S90, D4S43, and D4S125. It did not contain DNA from a centromeric STS in D4S10 or from QDPR (Fig. 4.16). In an effort to confirm that the 3.6 Mb product contained DNA from D4S10, a set of primers were designed from H5.52, a telomeric portion of D4S10 (43). A minimal amount of sequence information was determined from this region. Primers were designed to amplify a 550 bp fragment. PCR reactions using the H5.52 primers were performed on each fraction of the pulsed field gel used to resolve the 3.6 Mb cleavage product (85). The amplified products were resolved by gel electrophoresis and quantitated by DNA hybridization and phosphor imager analysis. As reported for the telomeric STS's, amplification was detected in fractions 6 and 7 with primers specific to H5.52 (Fig. 4.19) (85). This demonstrated that H5.52 was present in the 3.6 Mb fragment, and that the cleavage product contained the entire region from D4S10 to the telomere.

Site Specific Cleavage at a Target Site Telomeric to the HD Gene.

Although there is still a chance that HD is telomeric to all informative genetic markers and is located within the final 100 kb of chromosome 4 (54, 55), it is likely that the gene is between D4S10 and D4S113 (47, 60, 61). Approximately 1.5 Mb separates D4S113 from the telomere (50). While this extra DNA is not expected to interfere with efforts to map the HD mutation, a flanking cut at D4S113 would cleanly trim the desired fragment, and demonstrate the utility of this cleavage technique for genetic markers not located at the telomere.

Over 65 kb of sequence information surrounding D4S113 has been determined (86). This provided an ideal database to complete a computer

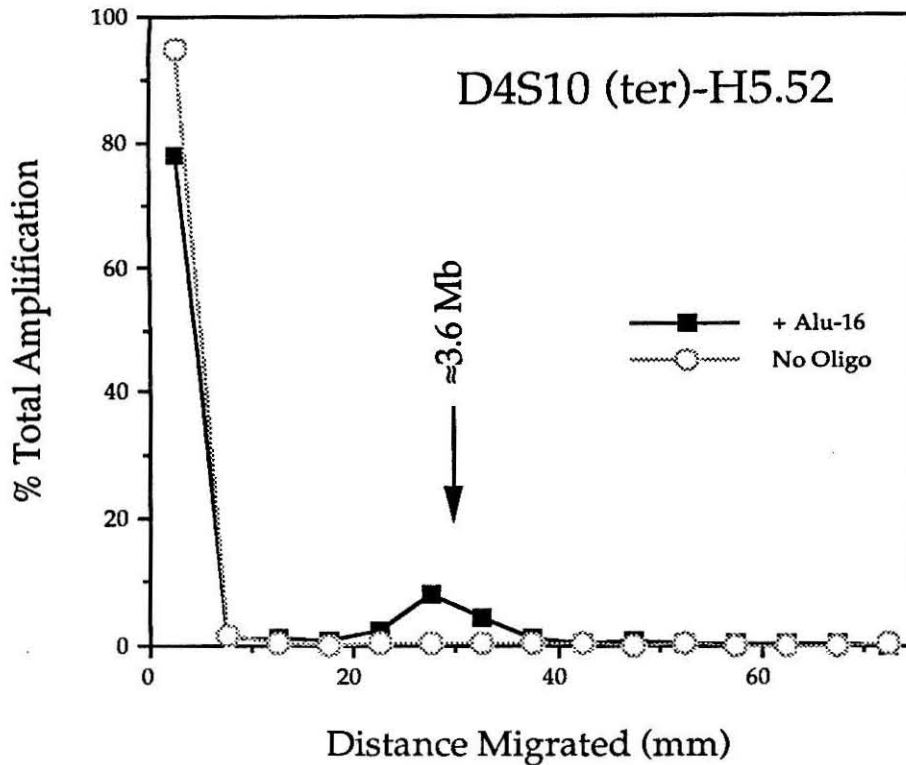


Fig. 4.19. PCR analysis of D4S10 (ter) locus H5.52 from lanes 3 and 4 gel Fig. 3A in Strobel, *et al. Science* 254, 1639-42 (1991). The no-oligonucleotide control lane (lane 3) is represented by gray lines and open circles. The reaction lane containing Alu-16 is represented by black lines and filled squares. Lanes 3 and 4 were cut into 15 sections 5 mm in size, placed in individual tubes, denatured for 5 min. at 96° C, vortexed, and 5 μ l used for PCR amplification with primers specific to H5.52. The primers used for amplification were: 1. 5' H5.52 (2): 5'-AAAGACGTGAGGTAGTGTCC-3'; and 2. 3' H5.52 (2): 5'-GTAGGA-AGCTGACACACCAG-3'. Amplification reactions were performed using the buffer and reagent conditions described in materials and methods. Amplification cycles were: 94° C 1.0 min., 58° C 1.0 min., 72° C 1 min. for 30 cycles. Amplification products were transferred to Nytran membrane and hybridized with radiolabeled with H5.52 radiolabeled DNA. The amplification intensity of amplification in each fraction was quantitated by phosphorimager analysis. The percent of total amplification, amplification in each fraction divided by total amplification in the lane, is plotted as a function of the distance migrated in the original pulsed field gel. Each square or circle corresponds to a 5 mm fraction. The peak observed in lane 4 is at the same position as observed for all other telomeric STS's. This corresponds to a migration distance of 25-35 mm and a size estimate of 3.6 ± 0.3 Mb. The presence of H5.52 in the 3.6 Mb fraction indicates the product contains all the DNA from D4S10 to the telomere.

search for sites targetable by triple helix mediated enzymatic cleavage. The analysis identified several potential target sites, of which two were selected and tested on plasmid DNA.

The first target selected was an 18 bp purine sequence that contained two T's and overlapped a HindIII site (Fig. 4.20 A). Three oligonucleotides containing T and MeC were synthesized for the target site with either G, D₃ (3-benzamidophenylamidazole), or D₈ (3-naphthamidophenylimidazole) opposite the pyrimidines (Fig. 4.20 B). HindIII restriction enzyme was reported to be inhibited by AluI methylation, although AluI methylase does not modify the same base as the cognate methylase of HindIII (78). Initial studies quickly demonstrated that AluI methylation only slowed cleavage by HindIII, but did not fully inhibit the activity of the enzyme. This enzyme combination would be insufficient to achieve site specific cleavage of human genomic DNA, as high fidelity endonuclease inhibition is essential.

HindIII methylase in low concentration and moderate purity was obtained from New England Biolabs to complete tests on the oligonucleotides (Fig. 4.20 C). Oligonucleotide HND-G mediated 70% target site cleavage at pH 6.6 (lane 2), but was highly pH sensitive and showed no protection above pH 6.8 (lane 3 and 4). Oligonucleotide HND-D₃ showed less pH sensitivity (87), mediating 95% cleavage at pH 6.6 (lane 5), and greater than 70% cleavage at pH 7.0 (lane 6). Little cleavage was observed at pH 7.4 (lane 7). HND-D₈ showed still improved resistance to pH (88), cleaving in high yield up to pH 7.0 (lanes 8 and 9) and at 25% cleavage at pH 7.4 (lane 9). However, HND-D₈ and other oligonucleotides containing the novel base D₈, showed moderate nonspecific inhibition of AluI methylase. Oligonucleotide HND-D₃ appeared to be the reagent of choice, but there was not sufficient HindIII methylase to conduct

Fig. 4. 20. Oligonucleotide mediated cleavage of D4S113 human chromosome 4 target site in pUCLEU2P plasmid DNA by HindIII methylase and endonuclease as a function of pH. **A.** Homopurine target site overlapping a HindIII restriction enzyme site. **B.** Oligonucleotides specific for target site in pUCLEU2P. They contain a G or novel bases D₃ and D₈ for recognition of the two TA base pairs in the target site. **C.** Enzymatic cleavage of pUCLEU2P DNA mediated by oligonucleotides HND-G, HND-D₃, or HND-D₈. Plasmid DNA (200 ng) was linearized with Sall and incubated with 1.0 μ M oligonucleotide in 1 mM spermine, 50 mM NaCl, 50 mM tris-HCl, pH 6.6-7.4, 10 mM EDTA, and 1 mM 2-mercaptoethanol for 12-14 hours at 22° C. S-adenosyl methionine (160 μ M), BSA (100 μ g/ml) and HindIII methylase (1 unit) were added to the reaction (final volume 20 μ l) and incubated at 22° for 2.0 hours. The DNA was phenol extracted, ethanol precipitated, 70% ethanol washed, dried, and resuspended in SstI restriction enzyme buffer containing 50 mM NaCl, 50 mM tris-HCl, 10 mM MgCl₂, pH 8.0 (25° C) and 20 units HindIII restriction enzyme. Digested the DNA to completion at 37° C for one hour. Resolved products on a 0.8% agarose gel in 1.0x TBE containing ethidium bromide. Visualized products by UV excitation. Intact linearized plasmid DNA is 6.6 kb and the desired cleavage products are 4.4 and 1.1 kb in size.

A.

HindIII/AluI

5'-CACAGAAGGATATAAAGAAAGCTTTAG-3'
 3'-GTGTCCTCCTATATTTCTTTCGAAATC-5'

B.

HND-G 5'- T^{Me}C T T^{Me}C MeC T G T G T T T^{Me}C T T T C-3'

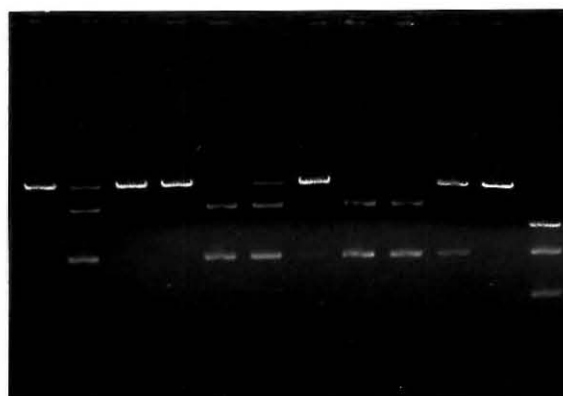
HND-D3 5'- T^{Me}C T T^{Me}C MeC T D₃ T D₃ T T T^{Me}C T T T C-3'

HND-D8 5'- T^{Me}C T T^{Me}C MeC T D₈ T D₈ T T T^{Me}C T T T C-3'

C.

	pH	7.0	6.6	7.0	7.4	6.6	7.0	7.4	6.6	7.0	7.4	7.0	7.0
HND-G		-	+	+	+	-	-	-	-	-	-	-	+
HND-D3		-	-	-	-	+	+	+	-	-	-	-	-
HND-D8		-	-	-	-	-	-	-	+	+	+	-	-
HindIII Methylase		-	+	+	+	+	+	+	+	+	+	+	-
HindIII Endonuclease		-	+	+	+	+	+	+	+	+	+	+	+
		1	2	3	4	5	6	7	8	9	10	11	12

6.6 kb →
 4.4 kb →
 2.2 kb →



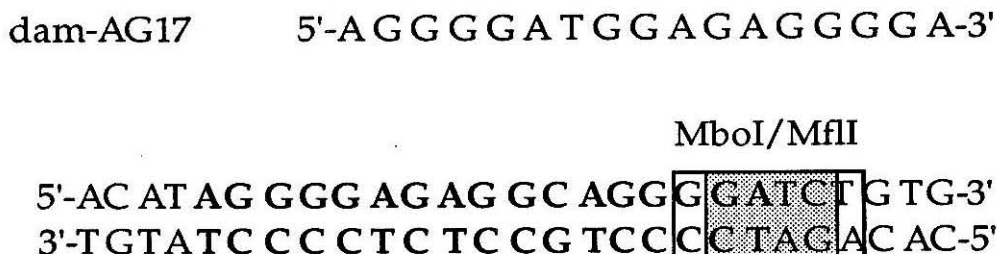


Fig. 4.21. Target site from D4S113 containing a 17 base pair G rich homopurine sequence overlapping the dam methylation site 5'-GATC-3'. The dam site is contained within an MfII recognition sequence, an endonuclease sensitive to dam methylation. MfII recognizes the sequence 5'-RGATCY-3'. The plasmid pUCLEU2P was constructed by inserting a duplex containing this sequence into the XhoI site of pUCLEU2. The sequence was designed to include a C instead of a T at the 3' end of the MfII site to generate a BamHI site. This was used to screen clones for the proper construction. Also shown is oligonucleotide dam-AG17, a purine oligonucleotide designed to bind the target duplex in the purine triple helix motif. The oligonucleotide uses T opposite the single CG base pair.

the analysis on yeast or human genomic DNA. The methylase must be purified before additional experiments can be conducted.

A second site within D4S113 was selected for cleavage. This site contained a 17 bp G rich purine sequence, interrupted by one C, and overlapping a dam methylase/MfII endonuclease site (Fig. 4.21). A duplex containing the target site was ligated into pUCLEU2 to generate pUCLEU2P. The high G content of the site made it untargetable by the pyrimidine triple helix motif previously used in this report. Instead, an oligonucleotide capable of binding in the purine motif was designed for the target site (89-91). The oligonucleotide contained G opposite each GC bp, A opposite each AT bp, and a T opposite the CG bp (92). Restriction enzyme competition experiments (93) demonstrated that the oligonucleotide completely saturated its target site at 1 μ M concentrations within one hour; however, no cleavage was observed if

the DNA was methylated and worked up using high pH for triplex disruption.

In previous experiments the pH sensitivity of the pyrimidine motif has been used to adjust the specificity of triple helix formation, and to displace the oligonucleotide following methylation (4, 68, 85). The purine motif is pH insensitive because there is no protonation requirement for binding (90). As a result the oligonucleotide was not displaced by the high pH and inhibited both the methylase and the restriction enzyme. While there is no documented low temperature procedure for removing the oligonucleotide from the binding site, Beal and Dervan have demonstrated that the purine motif is sensitive to the presence of polycations in solution, and that potassium inhibits triple helix formation possibly by stabilization of a four stranded competing structure between the oligonucleotides (90, 92). Attempts to remove the third strand by either of these procedures were unsuccessful. Addition of an oligonucleotide complementary to the third strand also failed to disrupt the triple helix. Thus, while the target site can be saturated with oligonucleotide, it remains to be demonstrated that the third strand can be displaced for restriction enzyme digestion.

Conclusion

Direct physical isolation of specific DNA segments from the human genome is an important objective in human genetics. Oligonucleotide-directed triple helix formation is a recognition motif for double helical DNA of sufficient generality and specificity to allow single site targeting of endogenous sequences in gigabase DNA. To test whether triple helix

mediated enzymatic cleavage was capable of liberating a specific segment of a human chromosome, we targeted the tip of human chromosome 4. Using a pool of partially degenerate pyrimidine oligonucleotides, a 16 bp purine sequence overlapping an AluI methylase/SstI endonuclease site was identified in a cosmid within a previously unsequenced locus proximal to the Huntington's disease gene. In the presence of an oligonucleotide specific for the target site, DNA isolated from a somatic cell hybrid containing human chromosome 4 and total mouse genomic DNA was exhaustively methylated. Subsequent triplex disruption and restriction enzyme digestion afforded in greater than 80% yield a single fragment, 3.6 Mb in size, containing the entire candidate region for the Huntington's disease gene. Each step in the process of target identification and cleavage is sufficiently generalizable that most cosmids should contain a sequence targetable by triple helix formation. This method may facilitate the orchestrated dissection of human chromosomes from normal and affected individuals into megabase sized fragments, and facilitate the isolation and analysis of candidate gene loci.

Materials and Methods.

Materials Restriction enzymes were obtained from New England BioLabs, Boehringer Mannheim, and Gibco BRL (SstI only). All methylases were obtained from New England Biolabs and used without further purification. T4 DNA ligase and the Klenow fragment of *E. coli* DNA polymerase were obtained from New England Biolabs. Sequenase 2.0 dideoxy

sequencing kits were purchased from United States Biolabs. Taq DNA polymerase and PCR buffers were obtained from Boehringer Mannheim. Perfect Match for PCR was obtained from Stratagene. PCR reactions were performed on a Perkin Elmer Cetus thermocycler. Prep-a-Gene DNA purification matrix and buffers were obtained from Bio-Rad. Large scale plasmid purification was performed using maxi-plasmid purification kits from Qiagen according to manufacture suggested protocols. Radioactive nucleotides ^{32}P - α -dCTP and ^{32}P - α -dATP were obtained from Amersham [10 mCi/ml, 3000 Ci/mmol]. Unlabeled mononucleotides were purchased from Pharmacia, mixed at 2.5 mM in each nucleotide, and stored in small aliquots for PCR reactions. Sephadex G-10 and G-50 matrices, spermine•4HCl (nuclease free), ampicillin, lauryl sarcosyl, 10 mg/ml salmon sperm DNA (fragmented and phenol extracted), and bovine serum albumin (20 mg/mL nuclease free) were obtained from Sigma. *Saccharomyces pombe* chromosomal DNA and concatenated λ DNA size standards and Incert low melting point (LMP) agarose were purchased from FMC Bioproducts. Electrophoresis grade agarose and LMP agarose were purchased from Gibco BRL. Plasmid pUC19 was obtained from Pharmacia. HD113.2B DNA prepared in LMP agarose was prepared by L. Riba in D. Housman's lab.

Oligonucleotides were synthesized on an Applied Biosystems 380 B DNA synthesizer. Cosmid and yeast chromosomal cleavage products were resolved on a CHEF-DRII pulsed-field gel electrophoresis (PFGE) chamber from Bio-Rad. Human chromosome 4 cleavage products were resolved on a CHEF MAPPER PFGE chamber from Bio-Rad. DNA blotting was accomplished using a Stratagene Pressure Blotter. 0.45 Micron Nytran membrane was obtained from Schleicher and Schuell. DNA nicking was accomplished with a UV Products short UV TS-20 transilluminator. The

DNA was immobilized by UV irradiation with a Stratagene Stratalinker. Radioactive emissions were detected on a Beckman LS 3801 liquid scintillation counter. DNA hybridization (Southern blotting) was accomplished in a Robins Scientific rotating hybridization incubator, Model 310. Autoradiography was performed using Kodak X-Omat AR film with an intensifying screen at -70°C and developed on a Kodak M35 A X-OMAT film processor. Densitometric traces were recorded on an LKB Bromma Ultrascan XL Laser Densitometer. Quantitative estimates of cleavage and PCR amplification were made on by storage screen phosphorescence using the Molecular Dynamics Phosphor Imager.

Oligonucleotide Synthesis and Purification. Oligonucleotides used for triple helix formation were deprotected in concentrated NH_4OH at 55°C for 12 to 24 hours in screw cap eppendorf tubes. Following deprotection the ammonia was removed with argon for 30-60 min. The oligonucleotides were dried by lyophilization and purified on a 20% acrylamide (45% urea) 1xTBE denaturing gel. The band was visualized by UV shadowing, cut from the gel, and eluted in 200 mM NaCl and 10 mM EDTA at 37°C overnight. The oligonucleotides were subsequently desalted by passage through a Sep-Pak reverse phase cartridge and eluted in 40% acetonitrile in water. The concentration was determined by UV absorbance and the oligonucleotide aliquoted into eppendorf tubes and dried for storage at -20°C .

Oligonucleotides for duplex insertion or PCR amplification were either not purified and used immediately after deprotection and neutralization, or in the case of a few oligonucleotides, purified by FPLC using a gradient of 100 mM triethyl ammonium acetate (TEAA) pH 7.0 and 40% acetonitrile in TEAA. The oligonucleotide was synthesized with the final dimethoxytrityl group attached at the 5' end, and the purification accomplished by separation

of the trityl-on material from all failure sequences. The trityl group was subsequently removed in 80% glacial acetic acid for 30 min., the solution dried, and the oligonucleotide desalted as described above.

Oligonucleotides containing the novel bases D3 and D8 were deprotected in 0.1 N NaOH at 55° C instead of concentrated ammonia. All other purification steps were as described both for gel and FPLC purification.

Cleavage Reactions. Complete protocols for the site specific cleavage of plasmid, cosmid, yeast, and somatic cell hybrid chromosomal DNA are included in the figure legends for the appropriate substrate. DNA hybridization was performed as described in Chapter 3.

Detection of Human Chromosome 4 Cleavage Products by the PCR. Human chromosome 4 cleavage products were detected by resolving the fragments on a 0.8% LMP agarose gel in 1x TAE at 2.0 V/cm, 14°C, 106° switch angle, 32 min. switch times for 72 hours. The full length of the reaction lanes containing the cleavage product were excised from the gel and cut into 5 mm sections perpendicular to the length of the gel. The DNA in each fraction was denatured by melting the agarose at 96° C for 5 min. and vortexed to evenly distribute the DNA throughout the volume of the agarose. PCR reactions were performed using 5 µl of the denatured solution in a total volume of 50 µl. Sequence tagged sites (STS) spanning the length of 4p16.3 (94) and an STS from QDPR that maps to 4p15.3 (95) were used for PCR analysis (66, 67). All amplification reactions were performed in 10 mM tris-HCl, pH 8.3 (at 25° C), 50 mM KCl, 1.5 mM MgCl₂, 0.001% (w/v) gelatin, 200 µM of each dNTP, 2 µM in each oligonucleotide primer, 0.25 units of Perfect Match, and 2.5 units of Taq DNA polymerase in 50 µl total volume.

D4S90 was the most telomeric STS analyzed by the PCR (44). The primers used for amplification were: D4S90-5', 5'-GTCCAGAGGAAGATG-

TGTAGG-GAC-3' and D4S90-3', 5'-CTACCACACCAGATCGACTAAGC-3' (94). Temperature conditions for amplification were 94° C 1.0 min., 60° C 1.0 min., and 72° C for 1.0 min. for 30 cycles. Amplification intensity was improved using 58° C hybridization temperature, but it also raised the background amplification.

D4S43 maps to the center of the HD candidate region (36). Primers specific for amplification of this locus were: D4S43-5', 5'-GACTGGTTG-TTTGAGGGCGTTG-3' and D4S43-3', 5'-TCCTTGACTCTGCTTCACC-3' (94). Amplification was accomplished using 28 cycles at 94° C 1.0 min., 60° C 1.0 min., and 72° C 1.0 min.

D4S125 is near the proximal boundary of the HD candidate region (48, 60). It is a highly polymorphic loci containing variable number of terminal repeats (VNTR). A region to the 3' side of a VNTR was chosen for PCR amplification. The primers were YNZ32-5', 5'-CTCTGGTCTGAGG-TGCTGAC-3' and YNZ32-3', 5'-CGGCAGCTGAGGAGGTGCCT-3' (48). Amplification was accomplished using 28 cycles at 94° C 1.0 min., 60° C 1.0 min., and 72° C 1.0 min. The amplification was rather weak and a very strong primer dimer was observed at a size just smaller than the desired product. A hot start might improve the signal with these primers (96).

An STS within D4S10 telomeric to the cleavage site was developed by partially sequencing H5.52 DNA (43). Amplification and detection of this STS is described in figure 4.15.

A centromeric STS within D4S10 was also tested to determine the nature of the 3.6 Mb cleavage product. The primers used were: D4S10-5', 5'-GGCACCTGGATCTCGGGC-3' and D4S10-3', 5'-GGAACGGGAGGCCAGC-3' (71). Amplification was accomplished using 30 cycles at 94° C 1.0 min., 60° C 1.0 min., and 72° C 1.0 min.

QDPR was the most centromeric STS analyzed by PCR. It maps one cytogenetic band below the expected cleavage site (95). Primers specific for this site were: QDPR-5', 5'-GTCACCTAACCTGTCTCAGTGTGG-3', and QDPR-3', 5'-GTAGTCAAGATGACAGCCACTGTC-3'. Amplification was accomplished using 30 cycles at 94° C 1.0 min., 60° C 1.0 min., and 72° C 1.0 min.

All amplification products were resolved on a 0.8% 1xTAE agarose gel and transferred to a Nytran membrane. The blot was hybridized as described in Chapter 3 with random primer radiolabeled product generated by amplification of the original clone. Amplification intensity was quantitated by phosphor imager analysis. The values were then standardized and plotted as a function of distance migrated verses percent total amplification in the lane.

References

1. H. E. Moser, P. B. Dervan, Sequence-Specific Cleavage of Double Helical DNA by Triple Helix Formation *Science* **238**, 645-650 (1987).
2. S. A. Strobel, H. E. Moser, P. B. Dervan, Double-Strand Cleavage of Genomic DNA at a Single Site by Triple Helix Formation *J. Amer. Chem. Soc.* **110**, 7927-7929 (1988).
3. S. A. Strobel, P. B. Dervan, Site-Specific Cleavage of a Yeast Chromosome by Oligonucleotide-Directed Triple-Helix Formation *Science* **249**, 73-75 (1990).
4. S. A. Strobel, P. B. Dervan, Single-site Enzymatic Cleavage of Yeast Genomic DNA Mediated by Triple Helix Formation *Nature* **350**, 172-174 (1991).

5. U. S. Department of Health and Human Services and the U. S. Department of Energy, in *The U.S. Genome Project: The First Five Years FY 1991-1995* (National Technical Information Service, Springfield, VA, 1990), pp. 11-14.
6. J. C. Stephens, M. L. Cavanaugh, M. I. Gradie, M. L. Mador, K. K. Kidd, Mapping the Human Genome: Current Status *Science* **250**, 237-244 (1990).
7. M. D. Adams *et al.*, Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project *Science* **252**, 1651-1656 (1991).
8. J. B. Martin, J. F. Gusella, Huntington's Disease: Pathogenesis and Management *New Eng. J. Med.* **315**, 1267-1276 (1986).
9. A. Reiner, R. L. Albin, K. D. Anderson, C. J. D'Amato, J. B. Penney, A. B. Young, Differential Loss of Striatal Projection Neurons in Huntington Disease *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5733-5737 (1988).
10. G. A. Graveland, R. S. Williams, M. DiFiglia, Evidence for Degenerative and Regenerative Changes in Neostriatal Spiny Neurons in Huntington's Disease *Science* **227**, 770-773 (1985).
11. N. S. Wexler *et al.*, Homozygotes for Huntington's Disease *Nature* **326**, 194-197 (1987).
12. R. H. Myers, J. Leavitt, L. A. Farrer, J. Jagadeesh, H. McFarlane, C. A. Mastromauro, R. J. Mark, *et al.*, Homozygote for Huntington Disease *Am. J. Hum. Genet.* **45**, 615-618 (1989).
13. C. Laird, Proposed Genetic Basis of Huntington's Disease *Trends Genet.* **6**, 242-247 (1990).
14. H. M. Slatkin, *Genetics* **40**, 246-251 (1955).
15. S. Henikoff, T. D. Dreesen, Trans-Inactivation of the *Drosophila* Brown Gene-Evidence for Transcriptional Repression and Somatic Pairing Dependence *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6704-6708 (1989).
16. R. H. Myers, J. J. Madden, J. L. Tiague, A. Falek, Factors Related to Onset Age of Huntington Disease *Am. J. Hum. Genet.* **34**, 481-488 (1982).
17. R. M. Ridley, C. D. Frith, T. J. Crow, P. M. Conneally, Anticipation in Huntington's Disease is Inherited Through the Male Line but May Originate in the Female *Am. J. Hum. Genet.* **25**, 589-595 (1988).

18. R. P. Erickson, Chromosomal Imprinting and the Parent Transmission Specific Variation in Expressivity of Huntington Disease *Am. J. Hum. Genet.* **37**, 827-829 (1985).
19. W. Reik, Genomic Imprinting: A Possible Mechanism for the Parental Origin Effect in Huntington's Chorea *J. Med. Genet.* **25**, 805-808 (1988).
20. J. F. Gusella *et al.*, DNA Markers for Nervous System Diseases *Science* **225**, 1320-1326 (1984).
21. J. Gusella, T. C. Gilliam, R. E. Tanzi, M. E. MacDonald, S. V. Cheng, M. Wallace, J. Haines, P. M. Conneally, N. S. Wexler, Molecular Genetics of Huntington's Disease *Cold Spring Harbor Quant. Biol.* **51**, 359-364 (1986).
22. J. F. Gusella, Location Cloning Strategy for Characterizing Genetic Defects in Huntington's Disease and Alzheimer's Disease *FASEB J.* **3**, 2036-2041 (1989).
23. C. Pritchard, D. R. Cox, R. M. Myers, The End in Sight for Huntington Disease? *Am. J. Hum. Genet.* **49**, 1-6 (1991).
24. J. F. Gusella *et al.*, A Polymorphic DNA Marker Genetically Linked to Huntington's Disease *Nature* **306**, 234-238 (1983).
25. P. S. Harper *et al.*, Genetic Linkage between Huntington's Disease and the DNA Polymorphism G8 in South Wales Families *J. Med. Genetics* **22**, 447-450 (1985).
26. S. E. Folstein *et al.*, Huntington's Disease: Two Families with Differing Clinical Features Show Linkage to the G8 Probe *Science* **229**, 776-779 (1985).
27. P. M. Conneally, J. L. Haines, R. E. Tanzi, N. S. Wexler, G. K. Penchaszadeh, P. S. Harper, S. E. Folstein, *et al.*, Huntington disease: No Evidence for Locus Heterogeneity *Genomics* **5**, 304-308 (1989).
28. J. F. Gusella, R. E. Tanzi, P. I. Bader, M. C. Phelan, R. Stevenson, M. R. Hayden, K. J. Hofman, A. G. Faryniarz, K. Gibbons, Deletion of Huntington's Disease-Linked G8 (D4S10) Locus in Wolf-Hirschhorn Syndrome *Nature* **318**, 75-78 (1985).
29. J. E. Landegent, N. J. in de Wal, Y. M. Fisser-Groen, E. Bakker, M. van der Ploeg, P. L. Pearson, Fine Mapping of the Huntington Disease Linked D4S10 Locus by Non-radioactive in situ Hybridization *Hum. Genet.* **73**, 354-357 (1986).

30. B. U. Zabel, S. L. Naylor, A. Y. Sakaguchi, J. F. Gusella, Mapping of the DNA Locus D4S10 and the Linked Huntington's Disease Gene to 4p16-p15 *Cytogenet. Cell. Genet.* **42**, 187-190 (1986).
31. T. C. Gilliam, *et al.*, Localization of the Huntington's Disease Gene to a Small Segment of Chromosome 4 Flanked by D4S10 and the Telomere *Cell* **50**, 565-571 (1987).
32. M. R. Hayden, C. Robbins, D. Allard, J. Haines, S. Fox, J. Wasmuth, M. Fahy, M. Bloch, Improved Predictive Testing for Huntington Disease by Using Three Linked DNA Markers *Am. J. Hum. Genet.* **43**, 689-694 (1988).
33. J. J. Wasmuth, L. R. Carlock, B. Smith, L. L. Immken, A Cell Hybrid and Recombinant DNA Library that Facilitate Identification of Polymorphic Loci in the Vicinity of the Huntington Disease Gene *Am. J. Hum. Genet.* **39**, 397-403 (1986).
34. T. C. Gilliam, S. T. Healey, M. E. MacDonald, G. D. Stewart, J. J. Wasmuth, R. E. Tanzi, J. C. Roy, J. F. Gusella, Isolation of Polymorphic DNA Fragments from Human Chromosome 4 *Nuc. Acids Res.* **15**, 1445-1458 (1987).
35. M. E. MacDonald *et al.*, A Somatic Cell Hybrid Panel for Localizing DNA Segments Near the Huntington's Disease Gene *Genomics* **1**, 29-34 (1987).
36. T. C. Gilliam *et al.*, A DNA Segment Encoding Two Genes Very Tightly Linked to Huntington's Disease *Science* **238**, 950-952 (1987).
37. J. J. Wasmuth, J. Hewitt, B. Smith, D. Allard, J. L. Haines, D. Skarecky, E. Partlow, M. R. Hayden, A Highly Polymorphic Locus Very Tightly Linked to the Huntington's Disease Gene *Nature* **332**, 734-736 (1988).
38. T. M. Pohl *et al.*, Construction of a NotI Linking Library and Isolation of New Markers Close to the Huntington's Disease Gene *Nuc. Acids Res.* **16**, 9185-9198 (1988).
39. B. Smith, D. Skarecky, U. Bengtsson, R. E. Magenis, N. Carpenter, J. J. Wasmuth, Isolation of DNA Markers in the Direction of the Huntington Disease Gene from the G8 Locus *Am. J. Hum. Genet.* **42**, 335-344 (1988).
40. M. R. Hayden *et al.*, A Polymorphic DNA Marker that Represents a Conserved Expressed Sequence in the Region of the Huntington Disease Gene *Am. J. Hum. Genet.* **42**, 125-131 (1988).

41. W. L. Whaley *et al.*, Mapping of D4S98/S114/S113 Confines the Huntington's Defect to a Reduced Physical Region at the Telomere of Chromosome 4 *Nuc. Acids Res.* **16**, 11769-11780 (1988).
42. J. E. Richards, T. C. Gilliam, J. L. Cole, M. L. Drumm, J. J. Wasmuth, J. F. Gusella, F. S. Collins, Chromosome Jumping from D4S10 (G8) toward the Huntington disease gene *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6437-6441 (1988).
43. M. I. Skraastad, E. Bakker, L. F. de Lange, M. Vegter-van der Vlis, E. G. Klein-Breteler, G. J. B. van Ommen, P. L. Pearson, Mapping of Recombinants Near the Huntington Disease Locus by Using G8 (D4S10) and Newly Isolated Markers in the D4S10 Region *Am. J. Hum. Genet.* **44**, 560-566 (1989).
44. S. Youngman *et al.*, A New DNA Marker (D4S90) is Located Terminally on the Short Arm of Chromosome 4, Close to the Huntington Disease Gene *Genomics* **5**, 802-809 (1989).
45. C. A. Pritchard, D. Casher, E. Uglum, D. R. Cox, R. M. Myers, Isolation and Field-Inversion Gel Electrophoresis Analysis of DNA Markers Located Close to the Huntington Disease Gene *Genomics* **4**, 408-418 (1989).
46. M. R. Altherr, B. Smith, M. E. MacDonald, L. Hall, J. J. Wasmuth, Isolation of a Novel Mildly Repetitive DNA Sequence that is Predominantly Located at the Terminus of the Short Arm of Chromosome 4 Near the Huntington Disease Gene *Genomics* **5**, 581-588 (1989).
47. W. L. Whaley *et al.*, Mapping of Cosmid Clones in Huntington's Disease Region of Chromosome 4 *Som. Cell Molec. Genet.* **17**, 83-91 (1991).
48. B. Richards, G. T. Horn, J. J. Merrill, K. W. Klinger, Characterization and Rapid Analysis of the Highly Polymorphic VNTR Locus D4S125 (YNZ32), Closely Linked to the Huntington Disease Gene *Genomics* **9**, 235-240 (1991).
49. L. A. Doucette-Stamm, L. Riba, B. Handelin, M. Difilippantonio, D. C. Ward, J. J. Wasmuth, J. F. Gusella, D. E. Housman, Generation and Characterization of Goss-Harris Hybrids of Human Chromosome 4 *Somatic Cell Mol. Genet.* In Press (1991).
50. M. Bucan *et al.*, Physical Maps of 4p16.3, the Area Expected to Contain the Huntington Disease Mutation *Genomics* **6**, 1-15 (1990).
51. M. E. MacDonald *et al.*, Recombination Events Suggest Potential Sites for the Huntington's Disease Gene *Neuron* **3**, 183-190 (1989).

52. B. A. Allitto *et al.*, Increased Recombination Adjacent to the Huntington Disease-Linked D4S10 Marker *Genomics* **9**, 104-112 (1991).
53. C. Robbins *et al.*, Evidence from Family Studies that the Gene Causing Huntington Disease is Telomeric to D4S95 and D4S90 *Am. J. Hum. Genet.* **44**, 422-425 (1989).
54. G. P. Bates *et al.*, A Yeast Artificial Chromosome Telomere Clone Spanning a Possible Location of the Huntington Disease Gene *Am. J. Hum. Genet.* **46**, 762-775 (1990).
55. C. Pritchard, D. Casher, L. Bull, D. R. Cox, R. M. Myers, A Cloned DNA Segment from the Telomeric Region of Human Chromosome 4p is not Detectably Rearranged in Huntington Disease Patients *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7309-7313 (1990).
56. W. R. A. Brown, P. J. MacKinnon, A. Villasante, N. Spurr, V. J. Buckle, M. Dobson, Structure and Polymorphism of Human Telomere-Associated DNA *Cell* **63**, 119-132 (1990).
57. R. G. Snell, L. Lazarou, S. Youngman, O. W. J. Quarrell, J. J. Wasmuth, D. J. Shaw, P. S. Harper, Linkage Disequilibrium in Huntington's Disease: An Improved Localization for the Gene *J. Med. Gene.* **26**, 673-675 (1989).
58. J. Theilmann *et al.*, Non-Random Association Between Alleles Detected at D4S95 and D4S98 and the Huntington's Disease Gene *J. Med. Gene.* **26**, 676-681 (1989).
59. S. Adam, J. Theilmann, K. Buetow, A. Hedrick, C. Collins, B. Weber, M. Huggins, M. Hayden, Linkage Disequilibrium and Modification of Risk for Huntington Disease *Am. J. Hum. Genet.* **48**, 595-603 (1991).
60. G. P. Bates, M. E. MacDonald, S. Baxendale, S. Youngman, C. Lin, W. L. Whaley, J. J. Wasmuth, J. F. Gusella, H. Lehrach, Defined Physical Limits of the Huntington's Disease Gene Candidate Region *Am. J. Hum. Genet.* **49**, 7-16 (1991).
61. M. E. MacDonald, C. Lin, L. Srinidhi, G. Bates, M. Altherr, W. L. Whaley, H. Lehrach, J. Wasmuth, J. F. Gusella, Complex Patterns of Linkage Disequilibrium in the Huntington Disease Region *Am. J. Hum. Genet.* **49**, 723-734 (1991).
62. B. S. Kerem, J. M. Rommens, J. A. Buchanan, D. Markiewicz, T. K. Cox, A. Chakravarti, M. Buchwald, *et al.*, Identification of the Cystic Fibrosis Gene: Genetic Analysis *Science* **245**, 1073-1080 (1989).

63. K. R. Thomas, M. R. Capecchi, Site-Directed Mutagenesis by Gene Targeting in Mouse Embryo-Derived Stem Cells *Cell* **51**, 503-512 (1987).
64. S. L. Mansour, K. R. Thomas, M. R. Capecchi, Disruption of the Proto-Oncogene *int-2* in Mouse Embryo-Derived Stem Cells: A General Strategy for Targeting Mutations to Nonselectable Genes *Nature* **336**, 348-352 (1988).
65. C. Denny, R. Baliga, S. A. Strobel, P. B. Dervan, *Unpublished Results* (1991).
66. R. K. Saiki, The Design and Optimization of the PCR in *PCR Technology* H. A. Erlich, Eds. (Stockton Press, New York, 1989), pp. 7-16.
67. H. A. Erlich, D. Gelfand, J. J. Sninsky, Recent Advances in the Polymerase Chain Reaction *Science* **252**, 1643-1651 (1991).
68. J. C. Hanvey, M. Shimizu, D. Wells, Site-Specific Inhibition of EcoRI Restriction/Modification Enzymes by a DNA Triple Helix *Nuc. Acids Res.* **18**, 157-161 (1990).
69. M. Koob, E. Grimes, W. Szybalski, Conferring Operator Specificity on Restriction Endonucleases *Science* **241**, 1084-1086 (1988).
70. M. Koob, W. Szybalski, Cleaving Yeast and *Escherichia coli* Genomes at a Single Site *Science* **250**, 271-273 (1990).
71. P. M. Stapleton, Sequence-Analysis of 3 Polymorphic Regions in the Human Genome Detected by the G8 Probe for RFLPS Associated with Huntingtons-Disease *Nuc. Acids Res.* **16**, 2735 (1988).
72. N. Kusukawa, T. Uemori, K. Asada, I. Kato, Rapid and Reliable Protocol for Direct Sequencing of Material Amplified by the Polymerase Chain Reaction *BioTechniques* **9**, 66-72 (1990).
73. M. Mihovilovic, J. E. Lee, An Efficient Method for Sequencing PCR Amplified DNA *BioTechniques* **7**, 14-16 (1989).
74. T. J. Povsic, P. B. Dervan, Triple Helix Formation by Oligonucleotides on DNA Extended to the Physiological pH Range *J. Am. Chem. Soc.* **111**, 3059-3061 (1989).
75. L. J. Maher, P. B. Dervan, B. Wold, Kinetic Analysis of Oligodeoxyribonucleotide-Directed Triple-Helix Formation on DNA *Biochemistry* **29**, 8820-8826 (1990).

76. G. A. Evans, G. M. Wahl, Cosmid Vectors for Genomic Walking and Rapid Restriction Mapping *Methods Enzymol.* **152**, 604-610 (1987).
77. G. M. Wahl, K. A. Lewis, J. C. Ruiz, B. Rothenberg, J. Zhao, G. A. Evans, Cosmid Vectors for Rapid Genomic Walking, Restriction Mapping, and Gene Transfer *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2160-2164 (1987).
78. C. Kessler, H. J. Holtke, Specificity of Restriction Endonucleases and Methylases *Gene* **47**, 1-153 (1986).
79. Y. Kariya, K. Kato, Y. Hayashizaki, S. Himeno, S. Tarui, K. Matsubara, Revision of Consensus Sequence of Human *Alu* Repeats-A Review *Gene* **53**, 1-10 (1987).
80. R. J. Britten, W. F. Baron, D. B. Stout, E. H. Davidson, Sources and Evolution of Human *Alu* Repeated Sequences *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4770-4774 (1988).
81. N. C. Dracopoli, M. M. Meisler, Mapping the Human Amylase Gene Cluster on the Proximal Short Arm of Chromosome 1 Using a Highly Informative (CA)_N Repeat *Genomics* **7**, 97-102 (1990).
82. L. C. Griffin, P. B. Dervan, Recognition of Thymine•Adenine Base Pairs by Guanine in a Pyrimidine Triple Helix Motif *Science* **245**, 967-971 (1989).
83. A. Andreadis, Y. Hsu, G. B. Kohlhaw, P. Schimmel, Nucleotide Sequence of Yeast *LEU2* Shows 5'-Noncoding Region has Sequences Cognate to Leucine *Cell* **31**, 319-325 (1982).
84. B. C. Hoopes, W. R. McClure, Studies on the Selectivity of DNA Precipitation by Spermine *Nuc. Acids Res.* **9**, 5493-5504 (1981).
85. S. A. Strobel, L. A. Doucette-Stamm, L. Riba, D. E. Housman, P. B. Dervan, Site-Specific Cleavage of Human Chromosome 4 Mediated by Triple Helix Formation *Science* **254**, 1639-1642 (1991).
86. C. Venter, *Unpublished Results* (1991).
87. L. C. Griffin, L. L. Kiessling, P. Gillespie, P. B. Dervan, Recognition of All Four Base Pairs of Double Helical DNA by Triple Helix Formation: Design of Nonnatural Deoxyribonucleosides for Pyrimidine•Purine Base Pair Binding *In Preparation* (1992).
88. L. L. Kiessling, P. Gillespie, P. B. Dervan, (In Preparation).

89. M. Cooney, G. Czernuszewicz, E. H. Postel, S. J. Flint, M. E. Hogan, Site-Specific Oligonucleotide Binding Represses Transcription of the Human *c-myc* Gene in Vitro *Science* **241**, 456-459 (1988).
90. P. A. Beal, P. B. Dervan, Second Structural Motif for Recognition of DNA by Oligonucleotide-Directed Triple Helix Formation *Science* **251**, 1360-1363 (1991).
91. D. S. Pilch, C. Levenson, R. H. Shafer, Structure, Stability, and Thermodynamics of a Short Intermolecular Purine-Purine-Pyrimidine Triple Helix *Biochemistry* **30**, 6081-6087 (1991).
92. P. A. Beal, Thesis (California Institute of Technology, 1994).
93. L. J. Maher, B. Wold, P. B. Dervan, Inhibition of DNA Binding Proteins by Oligonucleotide-Directed Triple Helix Formation *Science* **245**, 725-730 (1989).
94. J. F. Gusella *et al.*, (In Preparation).
95. S. Theune, J. Fung, S. Todd, A. Y. Sakaguchi, S. L. Naylor, PCR Primers for Human Chromosomes: Reagents for the Rapid Analysis of Somatic Cell Hybrids *Genomics* **9**, 511-516 (1991).
96. K. B. Mullis, The Polymerase Chain Reaction in an Anemic Mode: How to Avoid Cold Oligonucleotide-Ribonucleic Acid Fusion *PCR Methods Applic.* **1**, 1-4 (1991).

Chapter V

Proposal: Identification of the Huntington Disease Gene by Analysis of Cleavage Products

Triple helix mediated enzymatic cleavage is a highly specific technique for efficient cleavage of human chromosomes (1, 2). By targeting the chromosome 4p16.3 marker, D4S10, the Huntington Disease (HD) gene was liberated on a 3.6 Mb fragment resolvable by pulsed field gel electrophoresis (2). We propose to utilize this material for the physical identification of the HD gene. A pair of somatic cell hybrids have been created that contain chromosome 4 from affected and unaffected siblings (3). Despite the phenotypic difference between these siblings, the chromosomes are identical at all informative polymorphisms throughout the 4p16.3 region (4). The identification of a sequence difference between these matched chromosomes would be a strong candidate for the HD mutation. We propose to specifically cut both chromosomes at D4S10 and other distal genetic markers by triple helix mediated enzymatic cleavage, and in collaboration with Dr. David Housman, screen the purified telomeric material for the HD mutation. These screening strategies include: (i) analysis of the telomeric products for sequence variation by in-gel renaturation and differential S1 nuclease sensitivity (5, 6), and (ii) site specific cleavage of distal genetic markers on both chromosomes to generate 200-500 kb products for creation of a matched yeast artificial chromosome (YAC) library (7). These materials and techniques provide a strategy to physically search 4p16.3 from D4S10 to the telomere for the HD mutation.

Cleavage of Chromosome 4 at D4S10 and Determination of Purity.

D4S10 (8) was targeted for triple helix mediated enzymatic cleavage by the identification of a homopurine target sequence, 16 bp in length, partially overlapping an SstI site (5'-GAGCTC-3') (2). Methylation by AluI (5'-AGCT-3') in the presence of a 16 base oligonucleotide specific for the target site rendered all SstI sites resistant to digestion (9), except the single site left unmodified due to triple helix formation. Enzymatic treatment of DNA isolated from somatic cell hybrid HD113.2b containing human chromosome 4 (>200 Mb) and total mouse genomic DNA (\approx 10 gigabase pairs) (10) liberated a 3.6 Mb telomeric fragment in greater than 80% yield (2). Thus, site specific cleavage at D4S10 liberated a large, resolvable fragment that contained both HD candidate regions (4). We propose to use this technique to produce enriched telomeric material containing the HD candidate region and subsequently use it to identify the HD mutation.

Some issues about the cleavage product remain to be addressed. A fundamental issue is the relative purity of the isolated material. In HD113.2B, approximately 2% of total DNA is human, and only 0.04% of the DNA is from 4p16.3. The maximum cleavage yield in any reaction is 2 ng, if all the DNA is cut and all the cleaved product migrates into the gel. PCR suggests that at least half the DNA remains at the origin despite the high cleavage efficiencies. This reduces the yield to less than 1 ng per reaction. Ethidium bromide staining of reaction products resolved on pulsed field gels detects a faint smear that runs the length of the gel. Although the intensity of this band is light, it is significantly more DNA than the expected 1-2 ng that constitutes the maximum yield. Therefore, from the PCR and DNA hybridization analysis it is not clear to what extent the tip has been enriched.

Preliminary analysis of the reaction product by Alu-PCR (11) suggests that the telomeric fragment is significantly contaminated with other DNA. Alu-PCR amplifies only human genomic DNA, and each chromosome has a characteristic amplification pattern (11). HD113.2B contains only chromosome 4, and amplification results in six major amplification products (11). A second somatic cell hybrid, HHW693, contains human DNA from 4p15 to 4pter only (12). Alu-PCR of HHW693 generates four major amplification products, all distinct from the products generated from HD113.2B. Other regions in the chromosome dominate the amplification reaction when subsections of the chromosome are amplified. Alu-PCR was performed on fractions containing the cleavage product. Significant enrichment was expected to generate an amplification pattern akin to HHW693; however, the amplification pattern was more like that of HD113.2B. This suggests, but does not prove, that a significant portion of human chromosome 4 contaminates the cleavage product. If true, it implies that the DNA also contains significant mouse chromosomal DNA.

A slot blot assay could be used to get a better estimate of the purity of the cleavage product. Denatured DNA from each fraction on the gel could be transferred to a slot on a Nytran membrane and hybridized with either mouse or human repetitive sequences. Using standardized mixtures of total mouse and total human DNA, the relative ratios of mouse and human DNA in the gel fractions could be calculated. By determining the ratio of human to mouse DNA in the product fractions, compared to the ratio in the equivalent fractions from the lane that contained no oligonucleotide, the enrichment of 4p16.3 DNA could be calculated. The assay tolerates minimal cross-hybridization between mouse and human repetitive DNA, a stringency not obtained in a preliminary attempt to complete this analysis.

If the DNA is not sufficiently pure for subcloning and other types of analysis, the enrichment for 4p16.3 DNA could be improved by either (i) obtaining AluI methylase of higher specific activity and purity, or (ii) switching to the use of the cognate methylase for SstI endonuclease. The AluI methylase obtained from New England Biolabs was not highly purified, and did not retain consistent activity between lots when assayed for complete methylation of genomic DNA. Purification of the methylase, taking care to retain only the purest fractions and passing the enzyme over a S-homocysteine affinity column as the final step in purification, could result in methylase with greater specific activity and possibly improved product purity.

The use of SstI methylase might be the best approach to the preparation of pure product. SstI is not commercially available, and it is not known if SstI methylase is inhibited by triple helix formation. However, SstI methylase only modifies a small subset of the sites methylated by AluI, and might be an enzyme of greater specific activity for the D4S10 target site. SstI methylase has a six base pair recognition site which would eliminate methylation at statistically 16 fold fewer sites, as well as a vast number of sites in interspersed repetitive DNA (13, 14). This might improve the specific activity of the methylation reaction for the target site by 16 to 100 fold, and could significantly improve the purity of the resulting cleavage product. The development of an assay for DNA enrichment and improvements in the specific activity of the enzymes used for modification and cleavage are two critical areas that need to initially be addressed.

Cleavage of Matched Cell Lines. The generation of a telomeric fragment containing the HD candidate region is not equivalent to identification of the HD mutation, but the ability to generate enriched telomeric DNA from 4p16.3 would be useful toward that goal if it could be

generated from both an HD affected and a normal chromosome. Analysis could then be undertaken to explore the HD specific physical differences between these relatively short megabase fragments. Because polymorphisms would be extensive between two unrelated individuals, chromosomal substrates should be chosen that have maximal homozygosity through the region, while remaining phenotypically distinct relative to HD.

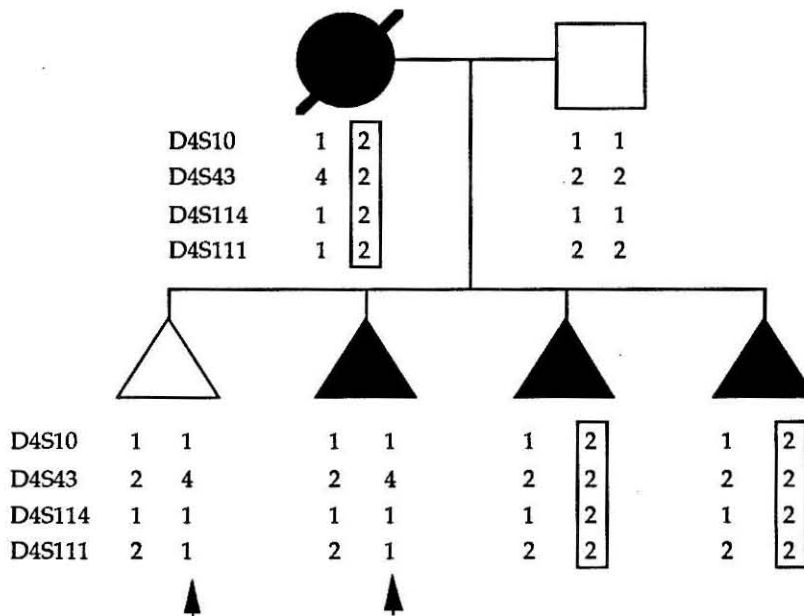


Fig. 5.1. Pedigree of family 1 from which matched chromosomes were derived. The affected and unaffected chromosomes indicated by arrows have been isolated in somatic cell hybrids HHW1248 and HHW1272, respectively.

Recombination in family 1 and two additional families (see Chapter 4) suggested a highly telomeric location for the HD gene because it segregated from all informative marker allele from D4S10 to D4S111, a locus approximately one megabase from the telomere (Fig. 5.1) (4, 15). However, telomere cloning and analysis of a highly telomeric restriction length polymorphism (RFLP) (100 kb) failed to detect the postulated recombination

(16, 17). Additional evidence, including telomere exchange (18) and linkage disequilibrium with internal loci (16, 17, 19, 20), suggested that an internal location for the HD gene is more likely (21, 22). This proximal location is supported by recombination in two other families (4, 23). If the HD gene is within the internal candidate region, then the critical recombination in families 1-3 must have been either local double recombination or gene conversion (23). If the double recombination that transferred the HD gene covered an interval of a few kilobase pairs or more, it would also exchange a number of polymorphisms between the two parent chromosomes. Therefore, sequence differences between the two chromosomes would either be the HD mutation or tightly linked to it. If a polymorphism between the two chromosomes could be identified, it would direct efforts toward the characterization of the HD mutation.

Triple helix mediated site specific cleavage of these two chromosomes could be used to prepare material for polymorphism analysis. John Wasmuth's group has generated a pair of somatic cell hybrids that contain maternal chromosome 4 from an affected and an unaffected sibling (Fig. 5.1), HHW1248 and HHW1272, respectively (3). PCR amplification and sequencing indicate that both of the chromosomes contain the complete SstI/triple helix target site within D4S10. Generation of 3.6 Mb fragments from these matched chromosomes will reduce the search for the HD gene to a comparison of two 3.6 megabase products containing a cluster of sequence differences localized to a narrow region. The experimental challenge becomes the development of a strategy to physically detect small differences between two large DNA fragments.

Mutational Analysis of Megabase Fragments. Using a modification of the in gel renaturation technique developed by Dr. Igor Roninson (5, 6, 24),

the DNA differences between these two cell lines can be identified. This technique was originally developed to analyze cell populations where gene amplification was suspected, but no probes corresponding to the amplified gene were available (5, 6, 24). The technique facilitated the detection of amplified fragments in restriction digests of total cellular DNA, identifying fragments that correspond to the essential gene within the amplicon.

The telomeric fragment isolated by triple helix mediated cleavage will be digested with a restriction enzyme with 6-bp specificity and labeled to high specific activity with T4 DNA polymerase (Fig. 5.2). This labeled fragment will serve as the tracer (5, 6, 24). Making use of the complete yeast artificial chromosomal (YAC) contig of 4p16.3 (25), YAC DNAs from the region will be digested with the same restriction enzyme and used as driver (DNA in excess) (5, 6, 24), to scan 4p16.3 for any mutations or physical differences between the two cell lines. The tracer and driver DNAs will be mixed together and run in a single lane of an agarose gel, followed by a series of in gel denaturations, renaturations, and S1 nuclease treatments (Fig. 5.2). In order to identify differences between these two cell lines, each YAC will be pooled with DNA cut from HHW1272 in one lane and HHW1248 in a second lane. After the in gel hybridization and nuclease treatment, the only fragments remaining in the gel will be those that are homologous between the telomeric fragment and the YAC (Fig. 5.2).

The second dimension of this procedure will involve two different gels. In the first case the nuclease treated lanes will be excised as an intact agarose strip, turned 90 degrees and loaded onto a second agarose gel. This will reveal a series of diagonal spots representing the DNA fragments present in the YAC used as driver. If any spots are off the diagonal, these represent differences between the cell line DNA and the YACs. However, if a difference

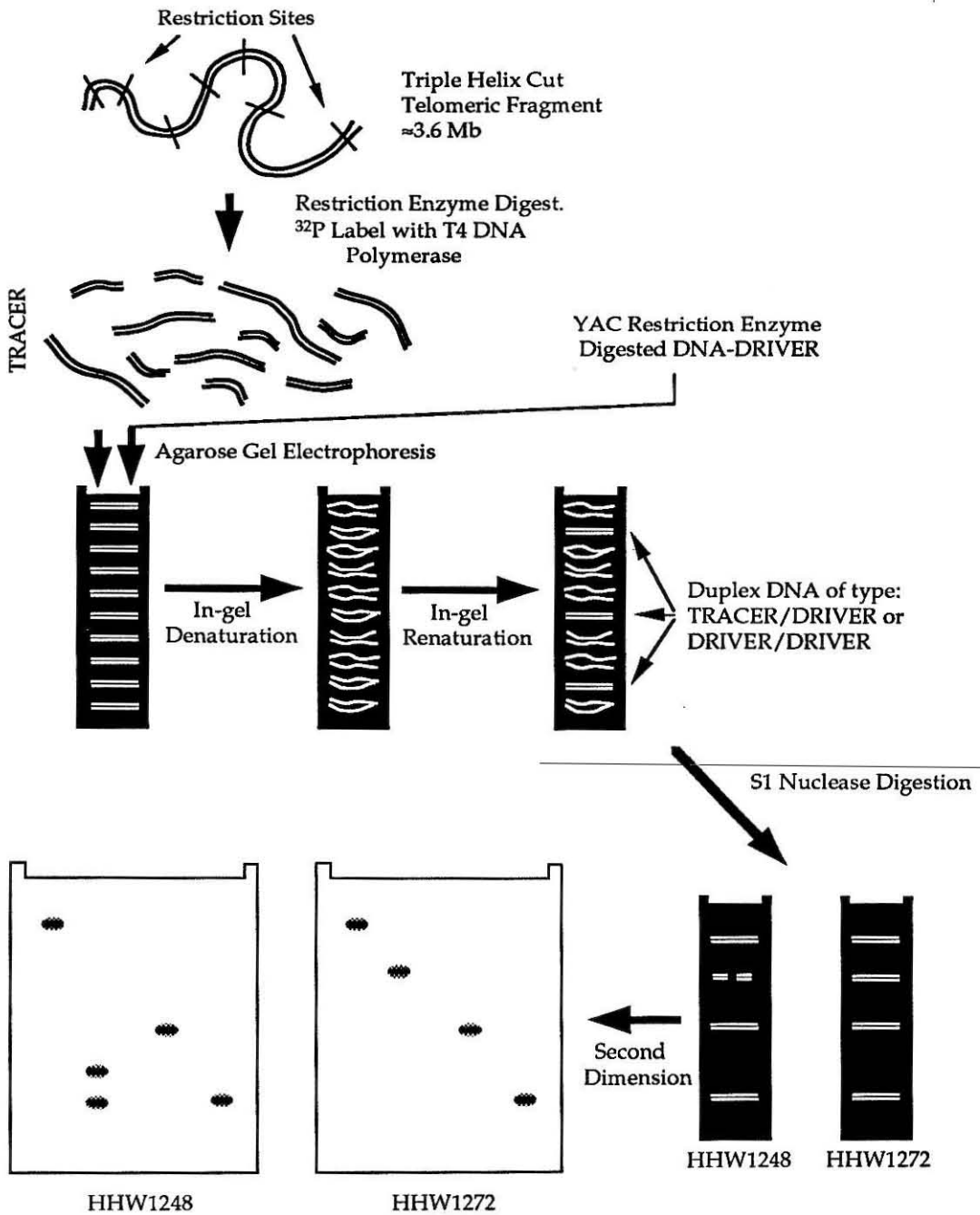


Fig. 5.2. Scheme for the identification of the HD or tightly linked mutation using differential in gel hybridization, S1 nuclease treatment, and two dimensional gel electrophoresis.

is only seen in the HD cell line, this represents a polymorphism between the HD and the normal chromosome. In the second class of gels proposed for detection of polymorphism, we will load the agarose strips on top of denaturing gradient gels to detect possible melting differences between the HD and normal chromosome hybrids with the YAC driver DNA.

If a difference between the two cell lines can be identified, we will narrow the location of the polymorphism by working with smaller YACs that map to the same interval, and eventually conducting the assay with cosmid DNA generated from the YAC inserts. Although the actual HD mutation might not be one readily detected by differential S1 reactivity, the theory of a double crossover or gene conversion implies that a number of other physically linked, but unrelated, mutations will be present in the region. The detection of any one alteration will define the region that underwent conversion. Once that region is identified to a single locus, sequencing will reveal the specific mutation..

Flanking Cut Sites and a Matching YAC Library. Linkage disequilibrium and two recombination events place the HD gene between D4S10 and D4S168 (E4) (21, 22). Cleavage at D4S168 would eliminate a large amount (≈ 1.5 Mb) of unnecessary DNA from the fragment to be searched. A flanking cut would reduce the fragment to approximately 2.0-2.5 Mb (21) and demonstrate the utility of this technique for the cleavage of loci not tightly linked to the telomere.

Craig Ventor has generated over 60 kb of sequence information from D4S113, a genetic marker just proximal to D4S168, but most likely telomeric to HD (26). A computer search of this sequence reveals a number of possible target sites for triple helix mediated enzymatic cleavage. These include, but are not limited to, an 18 bp purine rich sequence that overlaps a HindIII site

(5'-AAGCTT-3') and a G-rich sequence 17 bp in length that invades a dam methylase (5'-GATC-3') and MflI restriction enzyme site (5'-RGATCY-3'). Each site requires the development of a new enzyme system for genomic cleavage, but sites similar to these have been well behaved at the plasmid level.

If the material generated by cleavage and gel purification is insufficient in quantity or purity the region could be further subdivided into fragments of a size suitable for cloning into YAC vectors. This would require generating a cleavage event every 200 to 500 kb within the 2.5 Mb candidate region. Because a complete YAC contig is available for this critical region and much of the DNA is subcloned into cosmids (25), it should be a straightforward exercise to identify target sites that are separated by the desired interval.

If necessary, both the matched chromosomes will be cut at target sites spanning the HD candidate region to generate a series of fragment pairs with clonable ends. The products will be cloned into YAC's to produce a set of clones that should be identical except at the HD locus. This library of ordered clone pairs would be used in the Roninson gel experiments, and could greatly simplify the search for the desired mutation by making the DNA easily accessible in large quantity. However, any sequence difference identified would have to be confirmed within the original uncloned DNA to insure that the mutation was not generated by subcloning.

Both of these techniques provide a direct physical method for the identification of the mutation responsible for Huntington's disease. Previous analysis of HD has relied upon genetic mapping of random DNA fragments, physical mapping of these fragments, and linkage disequilibrium in HD families to map the HD gene. While these techniques have been invaluable to the localization of the mutation, the genetic data is sufficiently ambiguous

that they have not provided a definitive placement of the HD gene to one region within 4p16.3 (27). Efforts with these techniques have reached a level of diminishing return. The complexity of the problem requires that new strategies for the physical analysis of genomic DNA be developed (23). The site specific cleavage of human chromosome 4p16.3 into megabase and kilobase sized fragments provides a novel approach to identify the mutation responsible for Huntington's disease.

References

1. S. A. Strobel, P. B. Dervan, Single-site Enzymatic Cleavage of Yeast Genomic DNA Mediated by Triple Helix Formation *Nature* 350, 172-174 (1991).
2. S. A. Strobel, L. A. Doucette-Stamm, L. Riba, D. E. Housman, P. B. Dervan, Site-Specific Cleavage of Human Chromosome 4 Mediated by Triple Helix Formation *Science* 254, 1639-1642 (1991).
3. M. Altherr, J. Wasmuth, *Unpublished Results* (1991).
4. M. E. MacDonald, *et al.*, Recombination Events Suggest Potential Sites for the Huntington's Disease Gene *Neuron* 3, 183-190 (1989).
5. I. B. Roninson, Detection and Mapping of Homologous, Repeated and Amplified DNA Sequences by DNA Renaturation in Agarose Gels *Nuc. Acids Res.* 11, 5413-5431 (1983).
6. I. B. Roninson, H. T. Abelson, D. E. Housman, N. Howell, A. Varshavsky, Amplification of Specific DNA Sequences Correlates with Multi-drug Resistance in Chinese Hamster Cells *Nature* 309, 626-628 (1984).
7. D. T. Burke, G. F. Carle, M. V. Olson, Cloning of Large Segments of Exogenous DNA into Yeast by Means of Artificial Chromosome Vectors *Science* 236, 806-812 (1987).
8. J. F. Gusella *et al.*, A Polymorphic DNA Marker Genetically Linked to Huntington's Disease *Nature* 306, 234-238 (1983).

9. C. Kessler, H. J. Holtke, Specificity of Restriction Endonucleases and Methylases *Gene* **47**, 1-153 (1986).
10. L. A. Doucette-Stamm, L. Riba, B. Handelin, M. Difilippantonio, D. C. Ward, J. J. Wasmuth, J. F. Gusella, D. E. Housman, Generation and Characterization of Goss-Harris Hybrids of Human Chromosome 4 *Somatic Cell Mol. Genet.* In Press (1991).
11. D. L. Nelson, S. A. Ledbetter, L. Corbo, M. F. Victoria, R. Ramirez-Solis, T. D. Webster, D. H. Ledbetter, C. T. Caskey, *Alu* Polymerase Chain Reaction: A Method for Rapid Isolation of Human-Specific Sequences from Complex DNA Sources *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6686-6690 (1989).
12. J. J. Wasmuth, L. R. Carlock, B. Smith, L. L. Immken, A Cell Hybrid and Recombinant DNA Library that Facilitate Identification of Polymorphic Loci in the Vicinity of the Huntington Disease Gene *Am. J. Hum. Genet.* **39**, 397-403 (1986).
13. Y. Kariya, K. Kato, Y. Hayashizaki, S. Himeno, S. Tarui, K. Matsubara, Revision of Consensus Sequence of Human *Alu* Repeats-A Review *Gene* **53**, 1-10 (1987).
14. R. J. Britten, W. F. Baron, D. B. Stout, E. H. Davidson, Sources and Evolution of Human *Alu* Repeated Sequences *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4770-4774 (1988).
15. C. Robbins *et al.*, Evidence from Family Studies that the Gene Causing Huntington Disease is Telomeric to D4S95 and D4S90 *Am. J. Hum. Genet.* **44**, 422-425 (1989).
16. G. P. Bates *et al.*, A Yeast Artificial Chromosome Telomere Clone Spanning a Possible Location of the Huntington Disease Gene *Am. J. Hum. Genet.* **46**, 762-775 (1990).
17. C. Pritchard, D. Casher, L. Bull, D. R. Cox, R. M. Myers, A Cloned DNA Segment from the Telomeric Region of Human Chromosome 4p is not Detectably Rearranged in Huntington Disease Patients *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7309-7313 (1990).
18. W. R. A. Brown, P. J. MacKinnon, A. Villasante, N. Spurr, V. J. Buckle, M. Dobson, Structure and Polymorphism of Human Telomere-Associated DNA *Cell* **63**, 119-132 (1990).

19. R. G. Snell, L. Lazarou, S. Youngman, O. W. J. Quarrell, J. J. Wasmuth, D. J. Shaw, P. S. Harper, Linkage Disequilibrium in Huntington's Disease: An Improved Localization for the Gene *J. Med. Genet.* **26**, 673-675 (1989).
20. J. Theilmann *et al.*, Non-Random Association Between Alleles Detected at D4S95 and D4S98 and the Huntington's Disease Gene *J. Med. Genet.* **26**, 676-681 (1989).
21. G. P. Bates, M. E. MacDonald, S. Baxendale, S. Youngman, C. Lin, W. L. Whaley, J. J. Wasmuth, J. F. Gusella, H. Lehrach, Defined Physical Limits of the Huntington's Disease Gene Candidate Region *Am. J. Hum. Genet.* **49**, 7-16 (1991).
22. W. L. Whaley *et al.*, Mapping of Cosmid Clones in Huntington's Disease Region of Chromosome 4 *Som. Cell Molec. Genet.* **17**, 83-91 (1991).
23. M. E. MacDonald, C. Lin, L. Srinidhi, G. Bates, M. Altherr, W. L. Whaley, H. Lehrach, J. Wasmuth, J. F. Gusella, Complex Patterns of Linkage Disequilibrium in the Huntington Disease Region *Am. J. Hum. Genet.* **49**, 723-734 (1991).
24. P. Gros, J. Croop, I. Roninson, A. Varshovsky, D. E. Housman, Isolation and Characterization of DNA Sequences Amplified in Multidrug-Resistant Hamster Cells *Proc. Natl. Acad. Sci. U.S.A.* **83**, 337-341 (1986).
25. C. S. Lin *et al.*, New DNA Markers in the Huntington's Disease Gene Candidate Region *Somatic Cell Mol. Genet.* (In Press).
26. C. Venter, *Unpublished Results* (1991).
27. C. Pritchard, D. R. Cox, R. M. Myers, The End in Sight for Huntington Disease? *Am. J. Hum. Genet.* **49**, 1-6 (1991).

Appendix A

Oligonucleotide Sequences

Name	Sequence	Designed Application
λ /Adenovirus/Plasmid Affinity Cleaving		
λ -5'	5'-T*TTTCTTTTTTCTTTTCT-3'	λ Affinity Cleaving
λ -3'	5'-TTTTCTTTTTTCTTTTCT*-3'	λ Affinity Cleaving
λ -5'3'	5'-T*TTTCTTTTTTCTTTTCT*-3'	λ Affinity Cleaving
λ -5'Int	5'-T*TTTCT*TTTTTCTTTTCT-3'	λ Affinity Cleaving
λ -5'Int ²	5'-T*TTTCT*TTTTTCT*TTTCT-3'	λ Affinity Cleaving
λ -5'Int ² 3'	5'-T*TTTCT*TTTTTCT*TTTCT*-3'	λ Affinity Cleaving
λ -MeCBrU	5'-T*BBBMBBBBBBMBBBBBMB-3'	λ Affinity Cleaving
λ -18	5'-TTTTCTTTTTTCTTTTCT-3'	λ Competition
Ad-1	5'-T*CTTCTTCTTCTTCTTCCCTCC-3'	Adenovirus Cleavage
Ad-2	5'-T*TTTTTTTTTTTTTTTTTC-3'	Adenovirus Cleavage
Ad-3	5'-T*TCTCCTTTTCCTTCTC-3'	Adenovirus Cleavage
Ad-4	5'-T*TCTCCTTCTCTTCTTCTC-3'	Adenovirus Cleavage
Ad-5	5'-T*TCTTCTCCTTCTTCTTCTC-3'	Adenovirus Cleavage
p3'	5'-TTTTTCTCTCTCTCT*-3'	pDMAG10 Cleavage
p5'	5'-T*TTTTTCTCTCTCTCT-3'	pDMAG10 Cleavage
p5'3'	5'-T*TTTTTCTCTCTCTCT*-3'	pDMAG10 Cleavage
p16	5'-T*TTTTTCTCTCTCTCT-3'	pDMAG10 Mismatch
p17	5'-T*CTTTTTTCTCTCTCTCT-3'	pDMAG10 Mismatch
Cooperativity		
λ -9 3'A*	5'-TTCTTTTCT*-3'	Cooperativity
λ -9B	5'-TTTTCTTTT-3'	Cooperativity
λ -9C 5'T*	5'-T*TTTCTTTT-3'	Cooperativity
λ -7 3'A*	5'-CTTTTCT*-3'	Cooperativity
λ -11B	5'-TTTTCTTTTTT-3'	Cooperativity
λ -11M	5'-TTTTCTTTTTC-3'	Cooperativity
λ -11C5'T*	5'-T*TTTCTTTTTT-3'	Cooperativity
Coop-90	5'-TTCCTTTTT-3'	Cooperativity-Base Stacking
Coop-91	5'-TTTCCTTTT-3'	Cooperativity-Base Stacking
Coop-92	5'-TTTTCCTTT-3'	Cooperativity-Base Stacking

Plasmid/Yeast Construction		
AG-1	5'-TCGAGTCGAAGAAAAGAAGAG AAAGAAGAATTCGCGGCCGC-3'	Yeast Construct 1 (Not Used)
CT-1	5'-TCGAGCGGCCGCGAATTCCTTCT TTCTCTTCTTTCTTCGAC-3'	Yeast Construct 1 (Not Used)
AG-2	5'-TCGAGCCGGAAGAAAAGAAGAG AAAGAAGATCGATGCGGCCGC-3'	Yeast Construct 2 (Not Used)
CT-2	5'-TCGAGCGGCCGCATCGATCTTCTT TCTCTTCTTTCTTCCGGC-3'	Yeast Construct 2 (Not Used)
AAG	5'-TCGACTCGAAGAAAAGAAGAAAG AAAAAGAATTCG-3'	Yeast Construct A
ACT	5'-TCGACGAATTCCTTTCTTTCTTC TTTTCTTCGAG-3'	Yeast Construct A
BAG	5'-TCGACCCGGAAGAAAAGAAGAA AGAAAAAGGATCCG-3'	Yeast Construct B
BCT	5'-TCGACGGATCCTTTTTCTTTCTTCT TTTTCTCCGGG-3'	Yeast Construct B
Spe	5'-TCGACTTTTCTTTCCTTTTCTTTT ACTAGTAAAAGAAAAGGA AAGAAAAG-3'	Yeast Construct C
HindIII bistop	5'-TCGACTAGTAAAGAAAAGCTTTC TCCTTACTAG-3'	Yeast Construct D
HindIII bisbot	5'-TCGACTAGTAAAGAAGAAAGCTTTT TCTTTACTAG-3'	Yeast Construct D
SfiI Long	5'-GATCGGCCTCTAGGGCCCTGCA-3'	pUCSfi Construction
Sfi Short	5'-GGGCCCTAGAGGCC-3'	pUCSfi Construction
E-top	5'-TCGACTCGCTAATGGAAAGAGA GAGAGAGCTCACGTG-3'	Yeast Construct E
E-bottom	5'-TCGACACGTGAGCTCTCTCTTTT CCATTAGCGAG-3'	Yeast Construct E
113-Top	5'-TCGACACAGAAGGATATAAAGAA AGCTTTAG-3'	Yeast Construct H
113-Bot	5'-TCGACTAAAGCTTCTTTATATCC TTCTGTG-3'	Yeast Construct H
YST-Pur1	5'-TCGACATAGGGGAGAGGCAGGGG ATCCGTG-3'	Yeast Construct P
YST-Pur2	5'-TCGACACGGATCCCCTGCCTCTCC CCTATG-3'	Yeast Construct P

Sequencing/General PCR Primers		
LEU2 Sequence LEU2 PCR UP-MID A-MID Xho-235 Xho+250 UP 1200 UP 1201 UP-200 UP-380	5'-CCATTTTTAATAAGGCAATAAT-3' 5'-CAGTACCTTTAGCAAATTGTGGC-3' 5'-AAGGCTTACTTACTGATAGTAGA-3' 5'-ACATAACCACCCATAATGTAATAG-3' 5'-TCTATTACATTATGGGTG GTATGTT-3' 5'-GGTCAAGATATTTCTTGA ATCAGGC-3' 5'-GGTTTTCCAGTCACGACGT-3' 5'-AGGAAACAGCTATGACCATG-3' 5'-ATACCGCACAGATGCGTAAG-3' 5'-TCACATGTTCTTTCCTGCGT-3'	Yeast Insert Sequencing LEU2 PCR pUCLEU2 Sequencing pUCLEU2 Sequencing Yeast PCR Yeast PCR pUC19 Sequencing/PCR pUC19 Sequencing/PCR pUC19 PCR pUC19 PCR
Yeast Cleavage		
CT ^{T*} YST-T [*] MeCT ^{T*} MeCBrU ^{T*} YST-T [†] CTA MeCTA MeCBrU ^A CT ^B MeCT ^B MeCBrU ^B MeCT ^{BAD}	5'-T [*] TCTTTTCTTCTTTCTTTTT [*] -3' 5'-T [*] TCTTTTCTTCTTTCTTTTT [*] -3' 5'-T [*] TMTTTTMTTMTTMTTTTT [*] -3' 5'-T [*] BMBBBBMBBMBBBBMBBBT [*] -3' 5'-T [†] TCTTTTCTTCTTTCTTTTT [*] -3' 5'-CTTCTTTTCTTCTTTCTTTTTCTT-3' 5'-MTTMTTTTMTTMTT MTTTTTMTT-3' 5'-MBBMBBBBMBBMBBB MBBBBBMBB-3' 5'-CCTTCTTTTCTTCTTTCTTTTTCTT-3' 5'-MTTMTTTTMTTMTT TTTTMTT-3' 5'-MBBMBBBBMBBMBBB BBBBMBB-3' 5'-TMTTMTTTTMTTMTT TTTTMT-3'	Yeast Affinity Cleaving Yeast Affinity Cleaving Yeast Affinity Cleaving Yeast Affinity Cleaving Yeast Affinity Cleaving Yeast Enzymatic Cleavage Yeast Enzymatic Cleavage Yeast Enzymatic Cleavage Yeast Enzymatic Cleavage Yeast Enzymatic Cleavage Yeast Enzymatic Cleavage Yeast Enzymatic Cleavage Control
Yeast Marker PCR		
PGK.PCR1 PCK.PCR2 TRP3-A TRP3-B	5'-CATGTTTGACAGCTTATCAT-3' 5'-ATGTAAGTTTCACGAGGTTC-3' 5'-CCTTGAATCCCGACACATTG-3' 5'-TCTGCGTATTTGAGAGCCTG-3'	PGK1 PCR PGK1 PCR TRP3 PCR TRP3 PCR

Human Chromosome 4 Affinity Cleaving		
Hunt G-15	5'-T*MTTMTTTTGMGTTT-3'	HC4 Affinity Cleaving
Hunt G-24	5'-T*MTMTGGMTMTMTTTT MGMTTT-3'	HC4 Affinity Cleaving
Hunt Z-15	5'-TMTTMTTTTMD ₃ MTT-3'	HC4 Affinity Cleaving
Hunt Z-24	5'-TMTMTD ₃ D ₃ MTMTMTT TTMD ₃ MTT-3'	HC4 Affinity Cleaving
G81-6A2-G ₂	5'-T*TGMMTTTMTMTMTMT* -3'	HC4 Affinity Cleaving
G81-6A2-Z ₁	5'-T*TD ₃ MMTTTMTMTMTMT-3'	HC4 Affinity Cleaving
G81-6A2-Z ₂	5'-T*TD ₃ MMTTTMTMTMTMT*-3'	HC4 Affinity Cleaving
G83-5B1-G ₁	5'-T*MMMTGMMMTTMTTTT-3'	HC4 Affinity Cleaving
G83-5B1-Z ₁	5'-T*MMMTD ₃ MMMTTMTTTT-3'	HC4 Affinity Cleaving
G83-5B1-Z ₂	5'-T*MMMTD ₃ MMMTTMTTTT*-3'	HC4 Affinity Cleaving
G83-5B1-P ₁ Z ₁	5'-T*-P ₁ P ₁ P ₁ TD ₃ P ₁ P ₁ TTMTTTT-3'	HC4 Affinity Cleaving
G82-4C-151	5'-T*TCTTCTTTTCTTT-3'	HC4 Affinity Cleaving
G82-4C-152	5'-T*TCTTCTTTTCTTT*-3'	HC4 Affinity Cleaving
G82-4C-231	5'-T*TCTTCCCTTCTTTCTTTCTTT-3'	HC4 Affinity Cleaving
G82-4C-232	5'-T*TCTTCCCTTCTTTCTTTCTTT*-3'	HC4 Affinity Cleaving
Target Site Search by PCR		
TAG-PCR Eco-Seq ECO12PCR	5'-TACATGCCAATGACGACTGC-3' 5'-GGGTTTCCCAGTCACGAC-3' 5'-TACATGCCAATGACGACTGCGT GAATTCYYYYYYYYYYYY-3'	PCR Anchor Sequencing Primer EcoRI Purine Site Search (PCR)
DAM-TAG	5'-TACATGCCAATGACGACTGC RRRRRRRRRRRRRRRGATC-3'	dam Purine Site Search (PCR)
MSP-TAG	5'-TACATGCCAATGACGACTGC YYYYYYYYYYYYYYYYCCGG-3'	MspI Purine Site Search (PCR)
ECO-TAG	5'-TACATGCCAATGACGACTGC RRRRRRRRRRRRRRGAATTC-3'	EcoRI Purine Site Search (PCR)
TAQ-TAG	5'-TACATGCCAATGACGACTGC YYYYYYYYYYYYYYYYTCGA-3'	TaqI Purine Site Search (PCR)
ALU-TAG	5'-TACATGCCAATGACGACTGC RRRRRRRRRRRRRRRAGTC-3'	AluI Purine Site Search (PCR)
HAE-TAG	5'-TACATGCCAATGACGACTGC RRRRRRRRRRRRRRRGGCC-3'	HaeIII Purine Site Search (PCR)
PYR-TAG	5'-TACATGCCAATGACGACTGC YYYYYYYYYYYYYYYYYA-3'	Purine Site Search (PCR)
HPH-TAG	5'-TACATGCCAATGACGACTGC YYYYYYYYYYYYYYYYTCACC-3'	HphI Purine Site Search (PCR)
PST-TAG	5'-TACATGCCAATGACGACTGC YYYYYYYYYYYYYYYYCTGCAG-3'	PstI Purine Site Search (PCR)
CLA-TAG	5'-TACATGCCAATGACGACTGC YYYYYYYYYYYYYYYYATCGAT-3'	ClaI Purine Site Search (PCR)

AG-DAM16	5'-RRRRRRRRRRRRRRRRRGATC-3'	dam Purine Site Search No Anchor (PCR)
AG-ALU16	5'-RRRRRRRRRRRRRRRRRAGCT-3'	AluI Purine Site Search No Anchor (PCR)
AG-HAE16	5'-RRRRRRRRRRRRRRRRRGGCC-3'	HaeIII Purine Site Search No Anchor (PCR)
AG-15-ECO	5'-RRRRRRRRRRRRRRRRGAATTC-3'	EcoRI Purine Site Search No Anchor (PCR)
CT16-TAQ	5'-YYYYYYYYYYYYYYYYTCTGA-3'	TaqI Purine Site Search No Anchor (PCR)
CT16-MSP	5'-YYYYYYYYYYYYYYYYCCGG-3'	MspI Purine Site Search No Anchor (PCR)
CT-18	5'-YYYYYYYYYYYYYYYYYT-3'	Purine Site Search No Anchor (PCR)
Target Site Search by Achilles Heel (AC)		
Eco-Me dam-Me Msp-Me Hae-Me	5'-DDDDDDDDDDDDDDDDDDMTT-3' 5'-DDDDDDDDDDDDDDDDDDMT-3' 5'-MMDDDDDDDDDDDDDDDDDDT-3' 5'-DDDDDDDDDDDDDDDDDDMC-3'	EcoRI Purine Search (AC) dam Purine Search (AC) MspI Purine Search (AC) HaeIII Purine Search (AC)
Alu-Me Taq-Me Hha-Me Alu-Taq	5'-DDDDDDDDDDDDDDDDDDTTC-3' 5'-MTDDDDDDDDDDDDDDDDDDT-3' 5'-DDDDDDDDDDDDDDDDDDDC-3' 5'-GAPPPPPPPPPPPPPPPAG-3'	AluI Purine Search (AC) TaqI Purine Search (AC) HhaI Purine Search (AC) AluI or TaqI Purine Site Search (AC)
Hpa-Hae	5'-GGPPPPPPPPPPPPPPPPGG-3'	HpaII or HaeIII Purine Site Search (AC)
Human Chromosome 4 Enzymatic Cleavage		
Alu-16	5'-MMTTTMTMTMTMTMTC-3'	HC4 Enzymatic Cleavage (D4S10)
Alu-19G	5'-TTGMMTTTMTMTMTMTC-3'	HC4 Enzymatic Cleavage (D4S10)
Alu-GA16	5'-GAGAGAGAGAGAAAGG-3'	HC4 Enzymatic Cleavage (D4S10)
Alu-16P1	5'-P1P1TTTP1TP1TP1TP1TC-3'	HC4 Enzymatic Cleavage (D4S10)
HND- A62.8 (HND-G)	5'-TMTTMMTGTTTMTTTC-3'	HC4 Enzymatic Cleavage (D4S113-HindIII Site)
HND-D3	5'-TMTTMMTD3TD3TTTMTTTC-3'	HC4 Enzymatic Cleavage (D4S113-HindIII Site)
HND-D8	5'-TMTTMMTD8TD8TTTMTTTC-3'	HC4 Enzymatic Cleavage (D4S113-HindIII Site)
dam-17AG	5'-AGGGGATGGAGAGGGGA-3'	HC4 Enzymatic Cleavage (D4S113)

Human Chromosomal PCR Primers		
D4S10-5'	5'-GGCACCTGGATCTCGGGC-3'	HC4 D4S10 PCR
D4S10-3'	5'-GGAACGGGAGGCCAGC-3'	HC4 D4S10 PCR
SSG8-1	5'-CGGGCTGGGTTCCAGCAAGG-3'	HC4 D4S10 PCR
SSG8-4	5'-CATATGTTGAAGTCCTGCCG-3'	(Sequence Confirmation) HC4 D4S10 PCR
H5.52-3'	5'-GTCATCGCCTCTTTGTGGAT-3'	(Sequence Confirmation) HC4 D4S10 PCR
H5.52-5'	5'-CCCAGAGCTGTAGCTGCCAC-3'	(Did not Amplify) HC4 D4S10 PCR
5'-H5.52 (2)	5'-AAAGACGTGAGGTAGTGTCC-3'	(Did not Amplify) HC4 D4S10 PCR
5'-H5.52 (2)	5'-GTAGGAAGCTGACACACCAG-3'	HC4 D4S10 PCR
G83-0.4 3'	5'-AAGCTTGCATGCCTGCAGGT-3'	HC4 D4S10 PCR
G83-0.4 5'	5'-GGCTTCTGGCCTCCACAACT-3'	HC4 D4S10 PCR
YNZ32-5'	5'-CTCTGGTCTGAGGTGCTGAC-3'	HC4 D4S125 PCR
YNZ32-3'	5'-CGGCAGCTGAGGAGGTGCCT-3'	HC4 D4S125 PCR
D4S95-5'	5'-GAGTCTACCGGTGCCAAA-3'	HC4 D4S95 PCR
D4S95-3'	5'-TGGCCTCTCCAGATGGAATGT GCTC-3'	HC4 D4S95 PCR
D4S43-5'	5'-GACTGGTTGTTTGAGGGCGTTG-3'	HC4 D4S43 PCR
D4S43-3'	5'-TCCTTGACTCTGCTTCACC-3'	HC4 D4S43 PCR
D4S99-5'	5'-AAAGGTAGCAGTCCAGGC-3'	HC4 D4S99 PCR
D4S99-3'	5'-TGTGTGTCCCAGGCAG-3'	HC4 D4S99 PCR
D4S115-5'	5'-ATCAAGTCGAGGGACCTGGGCT-3'	HC4 D4S115 PCR
D4S115-3'	5'-CAGACAGCAGAGTCCACGGACAG-3'	HC4 D4S115 PCR
D4S97-5'	5'-GGTAGGGAGAGCTGCACGTG-3'	HC4 D4S97 PCR
D4S97-3'	5'-CAGACAGCAGAGTCCACGGACAG-3'	HC4 D4S97 PCR
D4S90-5'	5'-GTCCAGAGGAAGATGTG TAGGGAC-3'	HC4 D4S90 PCR
D4S90-3'	5'-CTACCACACCAGATCGACTAAGC-3'	HC4 QDPR PCR
QDPR-5'	5'-GTCACTAACCTGTCTCAGTGTGG-3'	HC4 QDPR PCR
QDPR-3'	5'-GTAGTCAAGATGACAGCCA CTGTC-3'	HC4 QDPR PCR
TC65	5'-TTGCAGTGAGCCAAGAT-3'	Alu-PCR
TC65R	5'-ATCTCAGCTCACTGCAA-3'	Alu-PCR
Enzymatic Cleavage at Misc. Sites		
myc-dam	5'-MMMMMTTTTMTTTTMT-3'	<i>c-myc</i> Enzymatic Cleavage
myc-Alu	5'-MTMTMTTMTTMTTTC-3'	<i>c-myc</i> Enzymatic Cleavage
EGFR-Eco	5'-TTMTTMTTTD ₃ MMTP ₁ P ₁ P ₁ T GMTT-3'	<i>EGFR</i> Enzymatic Cleavage
LIF-Alu	5'-MMGTTTMMTTTTC-3'	HC22 Enzymatic Cleavage

Triple Helix Specificity Assay (SELEX)		
DEG-17	5'-ACCTGCTGCAGLLLLLLLLLLLLLLLLL GAATTCGCAGTCGTCA TTGGCATGTA-3'	Triple Helix SELEX (Template)
64-mer	5'-TTGGCATCCTGCAGTCGAACTGAA TTCNNNNNNNNNNNNNNNNNNNGC TGAGCTCCCGTTCCGATGG-3'	Triple Helix SELEX (Template)
64-mer 3'	5'-CCATCGGAACGGGAGCTC-3'	Triple Helix SELEX (Primer)
64-mer 5'	5'-TTGGCATCCTGCAGTCGA-3'	Triple Helix SELEX (Primer)

Oligonucleotide sequences are listed with the following base abbreviations: T* (thymine-EDTA•Fe(II)), M (5-methylcytosine), B (5-bromouracil), P₁ (3-methyl-5-amino-1H-pyrazolo-[4,3-d]pyrimidine-7-one), D₃ (3-benzamido)phenylimidazole, D₈ (3-naphthamido)phenylimidazole, Y (50% T, 50% C), R (50% A, 50% G), L (40% G, 40% A, 10% T, 10% C), N (25% A, 25% G, 25% C, 25% T).

Nonenzymatic Double Strand Cleavage of a Yeast Chromosome at a Single Site in High Yield by Triple Helix Formation.

*Thomas J. Povsic, Scott A. Strobel and Peter B. Dervan**

*Arnold and Mabel Beckman Laboratories of Chemical Synthesis
California Institute of Technology
Pasadena, California 91125*

Abstract: *N*-Bromoacetyl oligodeoxyribonucleotides bind adjacent binding sites on double helical DNA by triple helix formation and alkylate single guanine positions on opposite strands in high yield. After depurination, the double strand cleavage yields in a plasmid (6 kbp in size) are greater than 85% and provide DNA fragments with termini that are ligatable with DNA fragments generated by restriction endonuclease digestion. The nonenzymatic double strand cleavage at a single site in a yeast chromosome, 340,000 base pairs in size, occurs in 85-90% yield.

Reliable models for the sequence-specific *recognition* of double helical DNA by low molecular weight peptides that bind in the minor groove,^{1,2} small protein-DNA binding domains,^{3,4} and oligonucleotide triple helix motifs⁵⁻⁷ now exist. This is due, in part, to a combination of footprinting/affinity cleaving methods¹ for determining sequence specificities, major and minor groove locations, and binding orientations of ligands on DNA, and the direct structural characterization of some of these complexes by nuclear magnetic resonance spectroscopy⁸ and x-ray diffraction analyses.⁹ The design of sequence-specific DNA *cleaving* molecules requires the integration of two separate functions, recognition and cleavage. It should be possible to combine DNA binding motifs with mechanism-based reactive functionalities capable of oxidation of the deoxyribose,^{10,11} electrophilic modification of the bases,¹² or hydrolysis of the phosphodiester backbone.¹³

Two criteria for successful bifunctional design are sequence specific reactions at designated single atoms within the bound complex and cleavage yields that are quantitative under physiological relevant conditions (neutral pH, 37 °C). In order to maximize stoichiometric reactions on DNA, the "cleaving functionality" must be sufficiently reactive to proceed at reasonable rates on DNA at 37 °C, be inert to aqueous media and buffer components, and not suffer unimolecular decomposition in competition with the desired reaction on DNA. In order to locate reactive moieties within angstroms of designated single atoms within a DNA binding site by designed synthetic molecules, information on groove location and relative base pair position within the ligand-DNA complexes is pivotal. In the absence of direct NMR or x-ray data, structural information can be provided by the affinity cleaving technique which relies on the generation of a *nonspecific diffusible* oxidant (hydroxyl radical) at a discrete location within a bound ligand-DNA complex by the

attachment of EDTA·Fe within a DNA binding motif.¹ Replacement of a *diffusible nonspecific* DNA cleaving moiety generated by EDTA·Fe(II) useful for studying DNA recognition (affinity cleaving) to a *nondiffusible base-specific* moiety^{11,12} is a key issue with respect to the design of sequence-specific DNA cleaving molecules. Sequence-dependent recognition can be coupled with sequence-dependent cleavage. One example is the conversion of distamycin-EDTA·Fe to *N*-bromoacetyldistamycin.¹² *N*-Bromoacetyldistamycin binds DNA in the minor groove at a five base pair A,T rich site and alkylates a single adenine position at N3 within those sites at 37 °C.¹²

Triple helix motif for DNA recognition. Pyrimidine oligodeoxyribonucleotide bind in the major groove of DNA parallel to the purine Watson-Crick strand through formation of specific Hoogsteen hydrogen bonds to the purine Watson-Crick base.^{5,6} Specificity is derived from thymine (T) recognition of adenine-thymine (AT) base pairs (TAT triplet); and N3-protonated cytosine (C⁺) recognition of guanine-cytosine (GC) base pairs (C⁺GC triplet). The generalizability of triple-helix formation has been extended beyond purine tracts to mixed sequences containing all four base pairs by the identification of other natural base triplets,^{5c,7} alternate strand triple-helix formation,^{5d} and designed nonnatural triple-helix motifs.¹⁴

Oligonucleotides equipped with EDTA·Fe can bind specifically to duplex DNA by triple-helix formation and produce double-strand cleavage at single sites in large DNA.⁵ Oligonucleotide-directed triple-helix recognition was shown to be a viable nonenzymatic approach for the site-specific cleavage of megabase DNA when an oligonucleotide-EDTA·Fe was targeted to a 20-base pair sequence in the 340-kilobase pair chromosome III of *Saccharomyces*

cerevisiae.¹⁵ Double-strand cleavage products of the correct size and location were observed, indicating that the oligonucleotide bound and cleaved the target site among almost 14 megabase pairs of DNA.¹⁵ Because oligonucleotide-directed triple-helix formation has the potential to be a general solution for DNA recognition, this result had implications for physical mapping of chromosomes.

The chemical yields for double strand cleavage by oligonucleotide-EDTA·Fe are typically low.⁵ In the case of plasmid (6 kb)^{5a} and λ phage (48.5 kb)^{5b} DNA, the yields for site-specific double strand cleavage are 15-25% respectively. This is adequate for the purposes of addressing questions related to molecular recognition, such as binding specificities, site size, and orientation. In the case of megabase DNA, the yields decline further 6%.¹⁵ If the full potential of oligonucleotide-directed recognition of double helical DNA is to be realized for site-specific cleavage of megabase DNA of high complexity, the design of oligonucleotides equipped with moieties capable of quantitative and base specific modification of DNA is an area for further development.

Sequence specific alkylation of double helical DNA by triple helix formation. We recently reported the design and synthesis of an oligodeoxyribonucleotide equipped with an *N*-bromoacetamide at the 5'-terminal thymidine that binds sequence specifically to a 19 base pair site in double-helical DNA by triple-helix formation and at 37 °C alkylates predominantly at a single guanine nucleotide adjacent to the target DNA sequence in high yield.^{16,17} Model building of triple-helical complexes indicated that pyrimidine oligodeoxyribonucleotides bound in the major groove parallel to the purine strand of the duplex and when equipped with a bromoacetyl moiety at the 5'-end, position the electrophile proximal to a guanine base two base pairs to the 5'-side of the target sequence. Reaction of

the electrophilic carbon with N-7 atom of guanine adjacent to the local triple helix results in covalent attachment of the oligonucleotide to the target sequence. Upon warming in the presence of base, depurination at the position of alkylation occurs and cleavage of one strand of the DNA backbone is observed. The DNA termini are 5' and 3' phosphoryl (Figure 1).¹⁸

Despite the high yields, *N*-bromoacetyl oligonucleotide-directed sequence specific alkylation of a single GC base pair in the major groove is limited to cleavage of one strand of the DNA binding site. The question arises whether this chemistry could be extended to cleavage of *both strands* of the double helical DNA target site. For example, two adjacent triple helix sites with two bromoacetyl oligonucleotides bound on opposite strands would afford alkylation/cleavage on both strands. This should result in DNA fragments with sequence specific overhangs suitable for ligation by restriction enzymes. If the overall yields are sufficiently high for double strand cleavage, this nonenzymatic approach to recognition and cleavage might be realized at single sites in megabase DNA.

We report here that two adjacent inverted oligonucleotide binding sites on double helical DNA, each containing a guanine residue two base pairs to the 5' side of each triple helix binding site affords double strand cleavage of DNA at a single site in high yield (Figure 2). From analysis of cleavage patterns on high resolution sequencing gels, the cleavage on each strand is at a single position. Analysis of a 6.7 kb plasmid revealed that double strand cleavage is achieved at a single site in 85% yield. Base work up results in DNA termini that are ligatable with complementary DNA fragments produced by restriction enzyme digestion. Within chromosomal DNA of high complexity, the nonenzymatic reaction is sufficiently efficient and specific that it produces

double strand cleavage at a single site in high yield within a yeast chromosome, 340,000 bp in size.

Results and Discussion

Synthesis of N-bromoacetyloligodeoxyribonucleotide (5) The synthesis of 5'-O-DMT-3'-phosphoramidite **3** follows in two steps from the previously described aminonucleoside **1**¹⁹ (Figure 3). Protection of the primary amine with flourenylmethyl chloroformate, followed by activation of the 3' hydroxyl with β -cyanoethyl-N,N-diisopropylchlorophosphoramidite²⁰ affords a thymidine derivative containing an appropriately protected linker-amine suitable for use in standard automated oligodeoxyribonucleotide synthesis. Incorporation at the 5' terminal base of an oligonucleotide was achieved in high yield. Deprotection of the amine occurs quantitatively under standard deprotection conditions (concentrated NH_4OH , 55 °C, 20 hr.), as analyzed by reverse-phase and anion-exchange HPLC, and polyacrylamide electrophoresis. Oligodeoxyribonucleotide-amine **4** was allowed to react with the N-hydroxysuccinimide of bromoacetic acid to afford N-bromoacetyloligodeoxyribonucleotide **5** which was purified by anion exchange HPLC.^{16,18}

Construction of target sequence. We previously demonstrated that efficient alkylation could be achieved at guanine two base pairs to the 5' side on the purine strand of a triple helical binding site using an N-bromoacetyloligonucleotide.¹⁶ By extension, adjacent inverted triple helix/alkylation sites would afford opposite strand alkylation and cleavage. Consideration of the cleavage patterns produced by triplex mediated alkylation indicates that a target sequence of the type 5'-

(pyrimidine)_mNC(N)_nGN-(purine)_m-3' which contains two inverted binding sites for triple helix formation may be suitable for double strand cleavage. This places cleavage positions on both strands of adjacent binding sites separated by n bases. Juxtaposition of the two inverted binding sites results in the formation of a local four-stranded structure at the junction of the binding sites (Figure 4). Therefore, the two sites of alkylation were separated by a few base pairs (n = 2).

The target site was designed with a single guanine two base pairs to the 5' side of each triplex binding site, the position most efficiently alkylated by the bound *N*-bromoacetyloligonucleotide 5. A single G flanked by A,T or C promotes cleavage specific to a single nucleotide position on each strand. Upon base workup 3, overhangs of size n + 1 are produced reminiscent of the type generated by several restriction enzymes. The double strand cleavage could result in DNA termini compatible for ligation with those produced via cleavage with a restriction endonuclease.

Based on the above considerations, the sequence 5'-CTTTTCCTTTCCTTTTCTTT TACTAGTAAAAGAAAAGGAAAGAAAAG-3' was cloned into the pUCLEU2 plasmid to generate pUCLEU2C. This sequence contains binding sites on each strand for the *N*-bromoacetyloligonucleotide 5 (5'-T₄MeCT₄MeC₂T₃MeCT₄-3') previously shown to efficiently modify the purine strand.¹⁶ After modification of the targeted guanine by the *N*-bromoacetyloligonucleotide, base treatment, depurination of guanine, and strand scission, the resulting 3'- overhang (CTA-3') is compatible with one of the ends (TAG-3') generated by Sfi I cleavage at the site (5'-GGCCTCTA/ GGGCC-3'). Finally, by cloning the site into a plasmid containing the *LEU2* gene,²¹ a gene essential for biosynthesis of leucine, it will be possible to specifically introduce this site into yeast

chromosome III. This will facilitate the study of the specificity of cleavage afforded by *N*-bromoacetyloligonucleotides from the single base pair to the megabase pair level.

High resolution analysis of opposite strand cleavage specificity. To insure that the oligonucleotide bound and efficiently modified the DNA solely at the targeted guanine bases, a 974 base pair *NarI/EcoRI* restriction fragment from pUCLEU2C containing the triple helix sites was radiolabeled with ^{32}P approximately 210 base pairs from the target sites. DNA labeled on either strand (3' or 5' end-labeled) was allowed to react with *N*-bromoacetyloligonucleotide 5 at 1 μM concentration at 37 °C for 36 hr. (0.8 mM $\text{Co}(\text{NH}_3)^{+3}$, 20 mM HEPES pH 7.4).^{16,22} After workup (0.1 M piperidine, 90 °C, 30 min.), analysis of the high resolution polyacrylamide gel reveals that modification occurred solely at a guanine nucleotide located two bases to the 5' side of each oligonucleotide binding site (Figure 5). From phosphorImager analyses, the cleavage yields were >96% on each strand. A plot of $\ln[\text{DNA}]_{\text{intact}}/[\text{DNA}]_{\text{total}}$ vs. time (pseudo-first order conditions) indicates the reaction between oligonucleotide 5 and double helical DNA is first order in DNA target concentration with pseudo-first order rate constants that correspond to a half life for alkylation within the triplex of 5.5 and 6.5 hr. for 3' and 5' end labeled strand, respectively.¹⁸ The measured rates are not significantly different from that reported for reaction of the *N*-bromoacetyloligonucleotide when only a single binding site is present.¹⁶

Double strand cleavage of plasmid DNA. To determine the yield of double strand cleavage, bromoacetyloligonucleotide 5 at 1 μM concentration (0.8 mM $\text{Co}(\text{NH}_3)_6^{+3}$, 20 mM HEPES pH 7.4) was allowed to react for 36 hr. at 37 °C with

a linearized plasmid DNA, 6.7 kbp in size. Cleavage at the adjacent triple helix target sites should produce DNA fragments 3.3 and 3.4 kilobase pairs in size which could be analyzed by nondenaturing gel electrophoresis. Treatment of the alkylated plasmid with 0.1% piperidine at 55 °C for 12 hr. in a buffer containing 10 mM EDTA, 10 mM Tris pH 8.0 and 100 mM NaCl resulted in two DNA fragments (Figure 6). Radiolabeled DNA was used to determine the efficiency of double helical cleavage. The ³²P end labeled products were separated by agarose gel electrophoresis and phosphorImager analyses indicated double helical cleavage efficiencies of 85%. No cleavage was observed in the absence of *N*-bromoacetyloligonucleotide 5.

Enzymatic ligation of DNA products resulting from double strand cleavage by *N*-bromoacetyloligonucleotides. Cleavage of pUCLEU2C with *N*-bromoacetyloligonucleotide 5 produced ends with the 3' overhanging sequence 5'-CTA-3' (Figure 7). End product analysis following piperidine workup indicated that the 3' and recessed 5' termini of both ends contain phosphate groups.¹⁸ Ligation reactions between restriction enzyme digested products involve joining of 5' phosphate to the 3' hydroxyl of complementary sequences. Digestion of the plasmid pUCSfiI with SfiI produced the 3' overhanging sequence, 5'-TAG-3', which is complementary to the end of the oligonucleotide digested product. Because the 3' end of the SfiI product contains a free hydroxyl group, it could be ligated to the 5' phosphate of the *N*-bromoacetyloligonucleotide digested DNA. It is unlikely that the 3' phosphate end of the oligonucleotide digested product is ligatable to the 5' phosphate of the restriction enzyme cut DNA; however, the remaining single strand nick could readily be repaired *in vivo* following transformation. The oligonucleotide cleavage products of pUCLEU2C (EcoRI cut) were ligated into the SfiI/EcoRI

fragment of pUCSfiI by T4 DNA ligase. The colonies were screened by α -complementation and five of each of the resultant plasmids were sequenced. All products were consistent with simple ligation of each of the *N*-bromoacetyloligonucleotide/EcoRI cut fragments into the *Sti*I/EcoRI sites of pUCSfiI. None of the products contained the G postulated to be lost upon modification and base workup. Thus, this synthetic cleavage agent produced DNA cleavage products that can be ligated into restriction enzyme digested DNA.

Double strand cleavage of a yeast chromosome 340 base pairs in size at a single site with *N*-bromoacetyloligonucleotides. Oligonucleotides equipped with the nonspecific DNA cleaving moiety EDTA·Fe(II) are capable of binding to and cleaving DNA at a single site within a yeast chromosome but in low yield. Moreover, unless conditions are carefully chosen, oligonucleotide-EDTA·Fe bind and cleave at partially homologous secondary sites. Coupling of a base specific cleaving moiety to an oligonucleotide capable of binding by triple helix formation may result in a molecule that may bind at secondary sites, but not react in the absence of a guanine residue two base pairs to the 5' side of the local triplex. Further specificity would be obtained by the requirement that triple helix sites be present on *both* strands of the DNA within a reasonable number (<20-30) of base pairs for double strand modification, cleavage and denaturation to occur.

To test the specificity of cleavage by *N*-bromoacetyloligonucleotides in megabase DNA, we constructed a yeast strain containing the double strand alkylation target site present in pUCLEU2C. The plasmid contains the target site adjacent to the *LEU2* gene, which encodes β -isopropylmalate

dehydrogenase, an enzyme that catalyzes the third step in leucine biosynthesis. A defect in this gene prohibits growth of yeast on leucine deficient minimal media. The PstI fragment from pUCLEU2 containing both the *LEU2* gene and the target sequence was homologously recombined into chromosome III of a *leu2* deficient yeast strain (SEY6210). Recombinants were selected by growth on minimal media lacking leucine. The target site was verified through use of the polymerase chain reaction (PCR) to amplify the target site insertion.²³ The target site is positioned on chromosome III such that cleavage should produce 230 and 110 kb products.

Chromosomal DNA from the appropriate strain, SEY6210C, was prepared in low melting point (LMP) agarose to prevent mechanical shearing of the DNA. The DNA was reacted with the N-bromoacetyloligonucleotide 5 using two cleavage cycles. This was accomplished by reacting the chromosomal DNA with oligonucleotide 5 as described above, followed by several washes in a low salt, high pH buffer to disrupt the triple helix and remove unreacted oligonucleotide from the solution. The DNA was reequilibrated in triple helix alkylation buffer, and a second aliquot of N-bromoacetyloligonucleotide 5 added. The DNA was then treated with 0.1% piperidine at 55 °C for 12 hr. to affect double strand cleavage. The cleavage products were resolved by pulsed field gel electrophoresis and visualized by ethidium bromide staining and by Southern blotting with the chromosome III specific marker, *HIS4*.²⁴

Incubation of total yeast chromosomal DNA with N-bromoacetyloligonucleotide 5 resulted in the cleavage of chromosome III to produce 230 and 110 kb products (Fig. 8). Ethidium bromide staining detected no other cleavage products within 14 megabase pairs of total yeast genomic

DNA, nor was there evidence of random degradation of total genomic DNA by the piperidine treatment. As a control, no cleavage products were detected on DNA from a yeast strain lacking a target site. No cleavage was observed in the absence of oligonucleotide. The cleavage efficiency was accurately quantitated by Southern blotting with the *HIS4* marker specific to the 110 kb product. A double strand cleavage efficiency of 90% was observed at the target site after two 20 hr. reaction cycles, and no low efficiency secondary cleavage sites were detected on chromosome III. Slightly reduced cleavage efficiency (85%) was detected if only one cleavage cycle was performed. These efficiencies for single site cleavage of chromosome III are in sharp contrast to oligonucleotide mediated affinity cleaving which cut at four secondary sites within chromosome III at a maximum overall yield of 6%. The specificity and efficiency of the nonenzymatic oligonucleotide-directed alkylation of DNA are comparable to that achieved by oligonucleotide mediated enzymatic cleavage.²⁵

Conclusion. We have examined oligonucleotide-directed alkylation of double helical DNA at inverted dimeric binding sites. The specificity of the cleavage reaction is derived from the requirement of two properly positioned oligonucleotide binding sites that each contained a G two base pairs to the 5' side of the triple helix site. This specified a total recognition/reaction site of 40 bp. The chemical cleavage reaction resulted in termini that were ligatable into compatible ends generated from restriction enzyme digestion. Thus, using a fully *nonenzymatic* approach to recognition and cleavage of single sites in megabase DNA, efficiencies comparable and specificities exceeding those attained by *enzymatic* cleavage of yeast genomic DNA mediated by triple helix formation were achieved.

Experimental

Collidine, fluorenylmethyl chloroformate and 2-cyano-*N,N*-diisopropylchlorophosphoramidite were purchased from Aldrich and used without further purification. Diisopropylethylamine (Aldrich) was distilled from CaH_2 prior to use. Dry solvents (DMF, methylene chloride, dioxane) were obtained from Fluka. Restriction enzymes were obtained from New England BioLabs, and used with the manufacturer's supplied buffers. Polynucleotide kinase, T4 DNA ligase and the Klenow fragment of DNA polymerase were obtained from Boehringer Mannheim. Radioactive nucleotides were purchased from Amersham. Prep-a-Gene matrix was purchased from Bio-Rad. Plasmids were purified using Maxi-plasmid purification kits from Qiagen. Dideoxy-sequencing kits were obtained from US Biochemicals. Yeast strain SEY6210 was obtained from Scott Emr. The HIS4 marker used from Southern blotting was the gift of Randy Schekman. Cobalt (III) hexamine trichloride was obtained from Kodak. Incert LMP agarose was obtained from FMC Biochemicals.

Deoxyribonucleoside 2. Aminonucleoside 1 (820 mgs., 1.27 mmoles lyophilized from dioxane) was dissolved in 5 ml of dry dioxane. Collidine (500 ml., 3.9 mmoles, 3 equivalents) was added, and the solution chilled to 0 °C. Fluorenylmethyl chloroformate (670 mgs., 2.6 mmoles, 2 equivalents, dissolved in 5 ml dry dioxane) was added over a ten minute period, and the reaction warmed to room temperature. After 20 min., tlc (5% MeOH in MeCl_2) showed one product. The solution was frozen and lyophilized. Chromatography in 2% MeOH in MeCl_2 afforded pure product (1.02 gms, 92% yield). ^1H NMR (CDCl_3): δ 7.72 (m, 2H), 7.55 (m, 2H). 7.13-7.39 (m, 14H), 6.80

(q, 4H), 6.35 (t, 1H, H1'), 4.48 (m, 1H, H3'), 4.37 (d, 2H), 4.15 (t, 1H), 4.04 (m, 1H, H4'), 3.74 (s, 6H, OCH₃), 3.38 (m, 2H, H5'), 3.06-3.20 (m, 4H), 2.37 (m, 2H, H2') 2.18 (m, 4H). IR (KBr) cm⁻¹: 3307, 3065, 2933, 1701, 1508, 1448, 1251, 1177, 1091, 1032, 909, 829, 759, 714. HRMS (FAB) calculated for C₅₀H₅₁O₁₀N₄ (M+H⁺): 867.3607. Found 867.3580.

Phosphoramidite 3. Deoxyribonucleoside 2 (150 mgs., 176 μmoles) was dissolved in 3 mls dry methylene chloride. Diisopropylethylamine (350 mls, 2 mmoles) and 2-Cyano-*N,N*-diisopropylchlorophosphoramidite (100 ml, 448 μmoles) were added, and the reaction stirred for 3 hr. The reaction was quenched with the addition of ethanol (0.5 ml) and diluted with ethyl acetate. The organic layer was washed with satd. NaHCO₃, water, and twice with satd. NaCl. After drying over Na₂SO₄ and concentration, flash chromatography using 500:15:5 methylene chloride:isopropanol:triethylamine solvent mixture afforded 110 mgs. of product (60% yield). ¹H NMR (CDCl₃): δ 8.40 (bs, 1H), 7.72 (t, J = 7.3 Hz, 2H), 7.55 (m, 3H), 7.13-7.39 (m, 13H), 6.79 (m, 4H), 6.33 (m, 1H, H1'), 5.49 (m, 1H), 4.59 (m, 1H, H3'), 4.39 (d, 2H-Fmoc, J = 6.6 Hz), 4.15 (m, 1H), 4.09 (m, 1H), 3.77 (s, 6H, OCH₃), 3.54 (m, 2H, H5'), 3.52-3.65 (m, 4H), 3.12-3.24 (m, 4H), 2.57 (t, J = 6.1 Hz, 1H), 2.37 (t, J = 6.1 Hz, 1H), 2.26 (m, 2H), 1.98-2.06 (m, 4H), 1.00-1.14 (m, 12H). IR (KBr) cm⁻¹: 3385, 3065, 2966, 2930, 1715, 1508, 1458, 1251, 1180, 1081, 1032, 979, 830, 760, 743, 700 ³¹P NMR (CDCl₃): δ 149.081, 148.812. HRMS calculated for C₅₉H₆₇N₆O₁₁P₁ (M·Na⁺): 1089.4536. Found 1089.4503.

Oligodeoxyribonucleotide-amine 4. Oligonucleotide 4 was synthesized on an Applied Biosystems 380B automated DNA synthesizer, and deprotected with concentrated NH₄OH at 55° for 20 hr. After lyophilization, the

oligonucleotide was purified by electrophoresis through a 20% denaturing polyacrylamide gel. The oligonucleotide was eluted from the excised band with 200 mM NaCl, 1 mM EDTA for 24 hr. at 37 °C, dialyzed against Millipore Ultra-pure water, and its concentration determined based on its UV absorbance (extinction coefficients were determined by summing the values for the individual nucleosides using values of $\epsilon_{260} = 8800$ for T and Talk, $\epsilon_{260} = 5700$ for MeC, $\epsilon_{280} = 6400$ for T and $\epsilon_{280} = 8300$ for MeC).

***N*-Bromoacetyloligonucleotide 5.** Ten nanomoles of the oligonucleotide were dissolved in 10 μ l of 200 mM borate buffer, pH 8.9. An equal volume of 250 mM *N*-hydroxysuccinimidylbromoacetate¹⁴ in DMF was added. After 10 min., the solution was injected onto a Hewlett-Packard 1090 HPLC equipped with a Vyadac oligonucleotide anion exchange 4.6 x 200 mm column. The probe was eluted in 20 mM Na₂HPO₄, 20% acetonitrile buffer, using a gradient of 40-80% NaCl over 60 min.. Elution times were 15.4 min. for the underivitized oligonucleotide 4, and 18.2 min. for the *N*-bromoacetyloligonucleotide 5. The product peak was collected and extracted with butanol to remove the acetonitrile. Salts were removed using a 3 ml Sephadex G-50-80 spin column, and the eluent precipitated with 3 volumes of NaOAc/ethanol and washed with 70% ethanol. The probe was diluted in water to an appropriate concentration and stored at -20 °C until needed.

Cloning of plasmids pUCLEU2C and pUCSfil. The plasmid pUCLEU2C was obtained by ligating a duplex of the inverted repeat oligonucleotide 5'-
T C G A C T T T T C T T T C C T -
TTTCTTTTACTAGTAAAAGAAAAGGAAAGAAAAG-3' into the XhoI site of plasmid pUCLEU2. The ligation mixture was digested with XhoI prior to

transformation to relinearize plasmid DNA lacking a duplex insert. Plasmid pUCSfi was obtained by forming a duplex from the oligonucleotides 5'-GATCGGCCTCTAGGGC-CCTGCA-3' and 5'-GGGCCCTAGAGGCC-3' and ligating into BamHI and PstI cut pUC19. The insert was designed to retain in frame translation of the *lacZ* gene so that α -complementation could subsequently be used for subcloning into the SfiI site. Transformation was performed using competent *E. coli* strain XL-1 blue (Stratagene). Plasmids were screened by digestion with SpeI (pUCLEU2C) or SfiI (pUCSfiI) and confirmed by dideoxy sequencing. Plasmids were purified from 500 ml cultures using a Qiagen maxi-purification column according to manufacturer's protocols.

Analysis at nucleotide resolution. pUCLEU2C was radiolabeled by first digesting with Nar I, to produce 5.6 and 1.1 kb fragments. The DNA was 5' end-labeled using Polynucleotide kinase and γ - ^{32}P ATP, and 3' end-labeled using the Klenow fragment to incorporate α - ^{32}P -dCTP and α - ^{32}P -dGTP. After labeling, unincorporated nucleotides were removed using a 1 ml Sephadex G-50-80 spin column. The DNA was digested with EcoR I, producing bands 4.5, 1.1, 9.0 and ~160 base pairs in size. The 1.0 kb band, containing the target sequence, was separated by 5% polyacrylamide gel electrophoresis, excised, and eluted at 37° using a solution of 200 mM NaCl, 1 mM EDTA. The DNA was purified by filtration through a 0.45 μm filter, butanol extraction and two ethanol precipitations.

Radiolabeled DNA was incubated in 20 μL reactions containing 0.8 mM $\text{Co}(\text{NH}_3)_6^{+3}$, 10 mM HEPES pH 7.4 and 100 nM *N*-bromoacetyl oligonucleotide 5. After 24 hr., the reaction was precipitated with the addition of NaOAc and

EtOH, 70% ethanol washed, and treated with 1.0% piperidine at 90 °C for 30 min. The samples were lyophilized, suspended in formamide buffer, and loaded onto a 6% denaturing polyacrylamide gel for analysis. Cleavage efficiencies were determined by phosphorImaging, using a model 400S phosphorImager and IQ software.

Rate constants for the reaction of *N*-bromoacetyloligonucleotide 5 with each strand were determined by running 10 parallel reactions, and quenching reactions by EtOH precipitation at desired time intervals. After piperidine treatment and gel electrophoresis, phosphorImaging was used to determine the fraction of intact DNA. Plotting $\ln(\text{fraction of intact DNA})$ vs. time yielded pseudo first-order plots from which rate constants were derived.¹²

Double stranded cleavage of plasmid DNA. Plasmid pUCLEU2C was linearized with Hind III and radiolabeled using Klenow and $-\alpha^{32}\text{P}$ dATP, resulting in a 6.7 kb DNA fragment labeled at both ends with the target site located approximately an equal distance from each end. The DNA was incubated in a solution (final volume 20 μl) containing 0.8 mM $\text{Co}(\text{NH}_3)_6^{+3}$, 10 mM Hepes pH 7.4, and 100 nM *N*-bromoacetyloligonucleotide for 24 hr. at 37°. The DNA was precipitated with NaOAc/ethanol, washed in 70% ethanol, dried, and treated with 100 μl of 0.1% piperidine, 100 mM NaCl, 10 mM Tris pH 7.4, 10 mM EDTA at 55° for 12 hr. The DNA was loaded onto a 0.8% ethidium containing agarose gel and electrophoresed for 5 hr. at 100 V. Cleavage efficiencies were determined by phosphorImaging.

Enzymatic ligation of double strand cleavage products. pUCLEU2C DNA was digested with EcoRI to completion and subjected to cleavage with *N*-

bromoacetyloligonucleotide 5. The resulting 450 and 600 bp fragments were gel purified and cleaned with Prep-a-Gene matrix according to manufacturer protocol. 5 µg of pUCSfi DNA was cut with SfiI and EcoRI, and the 2.8 kb product gel purified and cleaned as above. 10 ng of pUCSfi DNA in 10 µl T4 DNA ligase buffer (0.5 mM ATP, 20 mM Tris-HCl pH 7.4, 5mM MgCl₂, 5mM dithiothreitol, 50 µg/ml BSA) was added to 2 molar equivalents of either of the oligo/EcoRI fragments from pUCLEU2C, heated to 55 °C for 5 min. to denature all sticky ends, cooled to 16 °C, and allowed to react with 10 units T4 DNA ligase for 4 hr. The ligation mixture was used to transform competent *E. coli* (XL1-Blue, Stratagene) according to manufacturer's protocol and transformants selected on media containing ampicillin (50 mg/ml), and X-gal and IPTG for α-complementation. Growth overnight at 37 °C produced white colonies which were minipreped by SDS/NaOH lysis and sequenced by dideoxynucleotide chain termination using the USB Sequenase 2.0 kit.

Homologous recombination and production of alternate yeast strain.

Haploid yeast strain SEY6210 (*MAT α leu2-3 112 ura3-52 his 3- Δ 200 trp- Δ 901 ly s2-801 suc2- Δ 9 GAL*) was grown in a 200 ml liquid culture to an OD₆₀₀ of 0.7, harvested by centrifugation at 5,000 rpm for 5 min. at 50 °C, washed in sterile H₂O, centrifuged, resuspended in 40 mls 0.1 M LiO Ac, 10 mM Tris-HCl pH 8.0, 1 mM EDTA and incubated at 30 °C for 1 hr. Cells were again harvested by centrifugation, resuspended in 2 mls of LiOAc solution and 50 µl of competent cells aliquoted into sterile microcentrifuge tubes. 5 µg of PstI digested pUCLEU2C was added to the yeast solution and incubated at 30 °C for 1 hr. The cells were heat shocked for 5 min. at 37 °C and harvested by brief (3 s.) centrifugation. The supernatant was removed, the cells resuspended in 125 ml of sterile H₂O and plated on yeast minimal media plates lacking

leucine. Recombinants were detected after two days at 30 °C, selected colonies restreaked on a second minimal media plate, grown an additional two days at 30 °C, and screened for proper insertion of the oligonucleotide target sites adjacent to the *LEU2* gene by PCR amplification and DNA hybridization.

PCR amplification was performed using two oligonucleotides that hybridize approximately 250 bp on either side of the oligonucleotide insertion site (Xho-235 5'-TCTATTACATTATGGGTGGTATGTT-3' and Xho+250 5'-GGTCAAGATATTTCTTGAATCAGGC-3'). Chromosomal DNA was prepared from 0.5 mls of yeast liquid culture by glass bead lysis followed by PCR amplification with Taq DNA polymerase (Perkin Elmer Cetus) according to manufactures protocol. The 500 bp amplified products were digested with either XhoI or SpeI. Proper insertion of the oligonucleotide target sites would yield a PCR product cut only with SpeI, whereas an improperly modified strain would be XhoI cut. Intact chromosomal DNA from SpeI cut transformants was prepared in an agarose matrix,²⁶ separated on a pulsed-field gel and screened for unique insertion of the *LEU2* gene within chromosome III by Southern blotting with the random primer labeled EcoRI/KpnI fragment from pUCLEU2C. DNA from a yeast strain meeting these criterion was then used in alkylation reactions with the oligonucleotide.

Cleavage of yeast chromosomal DNA. DNA from yeast strains SEY6210 (no target site) and SEY6210C (+ target site) was isolated in low melting point agarose.²⁶ Agarose plugs containing yeast strain SEY6210C and SEY6210 were cut to approximately 2 mm thickness (volume approximately 80 μ l) and placed in 2 ml Eppendorf tubes. The plugs were washed 4 times with 900 μ l of 1 mM $\text{Co}(\text{NH}_3)_6^{+3}$, 20 mM Hepes pH 7.2, and 5 mM EDTA

(triplex/alkylation buffer). After removal of the last wash, the plugs were incubated with 120 μ l of triplex/alkylation buffer, *N*-bromoacetyloligonucleotide **5** was added to a final concentration of 1 μ M, and the plugs incubated at 37 °C. After 20 hr., the plugs were washed 4x 30 min. in 900 μ l of pH 9.5, 10 mM Tris, 10 mM EDTA, allowing the *N*-bromoacetyloligonucleotide to fully diffuse out of the agarose. The plugs were then washed 4 times for 15 min. each with 900 μ l of triplex/alkylation buffer, and incubated with 120 μ l triplex/alkylation buffer and *N*-bromoacetyloligonucleotide as before. After this second 20 hr. incubation, the agarose embedded DNA were washed 3 times in 900 μ l of 0.1% piperidine, 100 mM NaCl, 10 mM Tris, pH 8.0, 10 mM EDTA. After removal of the final wash, the DNA was heated in 900 μ l of the above piperidine solution at 55 °C for 12 hr. The plugs were transferred directly without melting to a 0.5x TBE 1.0% agarose gel and products separated by pulsed field gel electrophoresis using a BioRad Chef-DR™ II system. Electrophoresis was performed at 200 V for 24 hr., with switch times ramped from 10 to 40 s. over the first 18 hr., and from 60 to 90 s. over the last 6 hr. The gel was stained with ethidium bromide and photographed.

Southern blotting. The DNA was fragmented for efficient transfer by a 40 s. exposure of the gel with 254 nm UV transilluminator. The gel was soaked in 1.0 M NaCl, 0.5 M NaOH for 30 min. to denature the DNA, and then equilibrated in 1.5 M NaCl, 0.5 M Tris, pH 7.4 to neutralize the gel. The DNA was transferred using a Stratagene Pressure Control Station in 6xSSPE. *HIS4* hybridization was performed with a 250 bp *SalI/EcoRI* fragment derived from *HIS4-SUC2* fusion plasmid YCp503.²⁴ The fragment was labeled using random probe hybridization with degenerate 6mers and α -³²P-dCTP. The

membrane was prehybridized with 5 mls of a solution containing 6x SSPE, 10x Denhardt's, 1% SDS, 50 mg/ml Salmon sperm DNA solution at 42 °C for 2-4 hr. After removal of this solution, the membrane was washed with 2 mls of 6x SSPE, 50% formamide, 1%SDS, 50 mg/ml salmon sperm DNA solution at 42 °C for 30-60 min. The labeled probe was denatured by incubation at 37 °C for 5 min. with 1/10 volume of 0.1 M NaOH, and incubated with the membrane in 2 mls of 6x SSPE, 50% formamide, 1% SDS, 50 mg/ml salmon sperm DNA solution at 42 °C overnight. The blot was washed 4 times in 25 mls of 1x SSPE buffer, 1% SDS at 42 °C, exposed to film, and cleavage efficiencies quantitated using phosphorImaging.

Acknowledgement. We are grateful for generous support from the National Institutes of Health (GM-35724 and GM-43966), a Parsons predoctoral fellowship to TJP and a Howard Hughes Medical Institute predoctoral fellowship to SAS.

References

- (1) Dervan, P. B. *Science* **1986**, *332*, 464, and references cited therein.
- (2) (a) Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861. (b) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457. Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1985**, *107*, 5528. Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1986**, *108*, 5008. (c) Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574. (d) Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 6840. (e) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 7564.
- (3) (a) Sluka, J. P.; Bruist, M.; Horvath, S. J.; Simon, M. I.; Dervan, P. B. *Science* **1987**, *238*, 1129. (b) Sluka, J. P.; Griffin, J. H.; Mack, J. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 6369. (c) Oakley, M. G.; Dervan, P. B. *Science* **1990**, *248*, 847. (d) Sluka, J. P.; Horvath, S. J.; Glasgow, A. C.; Simon, M. I.; Dervan, P. B. *Biochemistry* **1990**, *29*, 6551. (e) Mack, D. P.; Sluka, J. P.; Shin, J. A.; Griffin, J. H.; Simon, M. I.; Dervan, P. B. *Biochemistry* **1990**, *29*, 6561. (f) Graham, K. S.; Dervan, P. B. *J. Biol. Chem.* **1990**, *265*, 16534.
- (4) (a) Steitz, T. A. *Quarterly Reviews of Biophysics* **1990**, *23*, 205-280. (b) Harrison, S. C.; Aggarwal, A. K. *Annual Review of Biochem.* **1990**, *59*, 933-969.
- (5) (a) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645. (b) Strobel, S. A.; Moser, H. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1988**, *110*, 7927. (c) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059. (d) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967. (e) Strobel, S. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 7286. (f) Maher, L. J.; Wold, B. J.; Dervan, P. B. *Science* **1989**, *245*, 725. (g) Horne, D. A.; Dervan, P.

- B. *J. Am. Chem. Soc.* **1990**, *112*, 2435. (h) Maher, L. J.; Dervan, P. B.; Wold, B. J. *Biochemistry* **1990**, *29*, 8820. (i) Plum, G. E.; Park, Y. W.; Singleton, S.; Dervan, P. B.; Breslauer, K. T. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 9436.
- (6) (a) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.L.; Thuong, N.T.; Lhomme, J.; Helene, C. *Nucleic Acids Res.* **1987**, *15*, 7749. (b) Praseuth, D.; Perrouault, L.; Le Doan, T.; Chassignol, M.; Thuong, N.T.; Lhomme, J.; Helene, C. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 1349. (c) Francois, J.-C.; Saison-Behmoaras, T.; Chassignol, M.; Thuong, N.T.; Helene, C. *J. Biol. Chem.* **1989** *264*, 5891. (d) Lyamichev, V.I.; Mirkin, S.M.; Frank-Kamenetskii, M.D.; Cantor, C.R. *Nucleic Acids Res.* **1988**, *16*, 2165. (e) Francois, J.C.; Saison-Behmoaras, T.; Thuong, N.T.; Helene, C. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 9198. (f) Perrouault, L.; Asseline, U.; Rivalle, C.; Thuong, N. T.; Bisagni, E.; Giovannangeli, C.; Le Doan, T.; Helene, C. *Nature* **1990**, *344*, 358-360. (g) Takasugi, M.; Guendouz, A.; Chassignol, M.; Decout, J. L.; Lhomme, J.; Thuong, N. T.; Helene, C. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5602-5606. (h) Mergny, J. L.; Sun, J. S.; Rougee, M.; Montenay-Garestier, T.; Barcelo, F.; Chomilier, J.; Helene, C. *Biochemistry* **1991**, *30*, 9791-9798.
- (7) (a) Cooney, M.; Czernuszewicz, G.; Postel, E.H.; Flint, S.J.; Hogan, M.E. *Science* **1988**, *241*, 456. (b) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360.
- (8) For example see: (a) Rajagopal, P.; Feigon, J. *Nature* **1989**, *239*, 637. (b) de los Santos, C.; Rosen, M.; Patel, D. *Biochemistry* **1989**, *28*, 7282. (c) Radhakrishnan, I.; de los Santos, C.; Patel, D. J. *J. Mol. Biol.* **1991**, *221*, 1408-1418. (d) Radhakrishnan, I.; de los Santos, C.; Live, D.; Patel, D. J. *Biochemistry* **1991**, *30*, 9022-9030.

- (9) (a) Wolberger, C.; Vershon, A. K.; Liu, B.; Johnson, A. D.; Pabo, C. O. *Cell* **1991**, *67*, 517-528. (b) Kissinger, C. R.; Liu, B.; Martin-Blanco, E.; Kornberg, T. B.; Pabo, C. O. *Cell* **1991**, *67*, 517-528. (c) Jordan, S. R.; Pabo, C. O. *Science* **1988**, *242*, 893-899. (d) Aggarwal, A. K.; Rodgers, D. W.; Drottar, M.; Ptashne, M.; Harrison, S. C. *Science* **1988**, *242*, 899-907.
- (10) (a) Hertzberg, R. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 313. (b) Hertzberg, R. P.; Dervan, P. B. *Biochemistry* **1984**, *23*, 3934.
- (11) (a) Mack, D. P.; Iverson, B. L.; Dervan, P. B. *J. Am. Chem. Soc.* **1988**, *110*, 7572. (b) Mack, D. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 4604.
- (12) (a) Baker, B. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1985**, *107*, 8266. (b) Dervan, P. B.; Baker, B. F. *Ann. N.Y. Acad. Sci.* **1986**, *472*, 51. (c) Baker, B. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 2700.
- (13) Basile, L. A.; Raphael, A. L.; Barton, J. K. *J. Am. Chem. Soc.* **1987**, *109*, 7550-7551.
- (14) (a) Koh, J. S.; Dervan, P. B. *J. Am. Chem. Soc.*, in press. (b) Kiessling, L. L.; Griffin, L. C.; Dervan, P. B. *Biochemistry*, in press.
- (15) Strobel, S. A.; Dervan, P. B. *Science* **1990**, *249*, 73.
- (16) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 9428.
- (17) For sequence-specific alkylation of double helical DNA see: (a) Fedorova, O. S.; Knorre, D. G.; Podust, L. M.; Zarytova, V. F. *FEBS Lett* **1988**, *228*, 273. (b) Vlassov, V. V.; Gaidamakov, S. A.; Zarytova, V. F.; Knorre, D. G.; Levina, A. S.; Nekona, A. A.; Podust, L. M.; Fedorova, O. A. *Gene* **1988**, *72*, 313. (c) Shaw, J. P.; Milligan, J. F.; Krawczyk, S. H.; Matteucci, M. *J. Am. Chem. Soc.* **1991**, *113*, 7765-7766.
- (18) T. Povsic Ph.D. thesis, California Institute of Technology 1991.

- (19) Dreyer, G. B.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 968.
- (20) M. J. Gait, Ed. *Oligonucleotide Synthesis: a Practical Approach*; IRL Press: Oxford, 1984.
- (21) Broach, J. R. et al. *Gene* **1979**, *8*, 121. Andrea-dis, A.; Hsu, Y. P.; Kohlhaw, G. B.; Schimmel, P. *Cell* **1982**, *31*, 319. Fischhoff, D. A.; Waterston, R. H.; Olson, M. V. *Gene* **1984**, *27*, 239.
- (22) It is known that triple helix formation is facilitated by the addition of certain cations.^{5,6} However, the presence of spermine and Tris buffers reduced the cleavage reaction yields, presumably due to reaction with the bromoacetamide moiety; therefore, the polycation $\text{Co}(\text{NH}_3)^{+3}$ and Hepes buffer were substituted.¹⁶ In control experiments, we observed no significant reaction between $\text{Co}(\text{NH}_3)^{+3}$ and the bromoacetyloligonucleotide (at 37 °C).
- (23) A 485 bp segment of chromosome III containing the target insertion site was amplified using flanking PCR primers. The resulting product was digested with SpeI, a restriction enzyme indicative of the presence of the target site, and with XhoI, a restriction enzyme indicative of the wild type sequence. The PCR product was cut only by SpeI which suggested that the desired construct had been achieved. Southern blotting with the KpnI-EcoRI fragment from the *LEU2* gene verified that the recombination had occurred on chromosome III, and not on another chromosome in the yeast genome.
- (24) (a) Keeseey, J. K.; Bigelis, R.; Fink, G. R. *J. Biol. Chem.* **1979**, *254*, 7427. (b) Deshaies, R. J.; Schekman, R. *J. Cell Biol.* **1987**, *105*, 633.
- (25) Strobel, S. A.; Dervan, P. B. *Nature* **1991**, *350*, 172.
- (26) Strobel, S. A.; Dervan, P. B. *Methods in Enzymology*, in press.
- (27) Iverson, B. L.; Dervan, P. B. *Nucleic Acids Res.* **1987**, *15*, 7823.

- (28) Maxam, A. M.; Gilbert, W. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 560.

Figure Legends

Fig. 1. The *N*-bromoacetamide electrophile is localized in the major groove by triple helix formation proximal to a guanine (G) in the Watson-Crick duplex target site. Alkylation at N7 of G followed by depurination results in backbone cleavage to afford 5' and 3' phosphate termini.

Fig. 2. Oligodeoxyribonucleotides bind to adjacent inverted binding sites on double helical DNA by triple helix formation and alkylate/depurinate at single guanine positions on opposite strands.

Fig. 3. Scheme for construction of Fmoc-protected thymidine 2-cyanoethyl-*N,N*-diisopropylphosphoramidite **3** used in the synthesis of oligonucleotide **4**. Modification of oligonucleotide **4** with *N*-hydroxysuccinimidyl bromoacetate afforded the bromoacetyloligonucleotide **5**.

Fig. 4. Ribbon model of *N*-bromoacetyloligonucleotide-directed double strand cleavage of DNA by triple helix formation at inverted adjacent half sites.

Fig. 5. (left) Autoradiogram of a high-resolution 6% denaturing polyacrylamide gel of cleavage products from reaction of *N*-bromoacetyloligonucleotide **5** with ³²P 3'- or 5'-end-labeled 0.9 kbp restriction fragment from plasmid pUCLEU2C. Lanes 1-4 are 5'-end-labeled DNA. Lanes 5-8 contain 3' end-labeled DNA. Lanes 1 and 8 are

A-specific chemical sequencing reactions for the 5' and 3' end-labeled fragments. Lanes 2 and 7 are G-specific chemical sequencing reactions for the 5' and 3' end-labeled fragments.²⁸ Lanes 3 and 6 contain DNA incubated for 36 hr. in the absence of 5'-*N*-bromoacetyl oligonucleotide 5, followed by piperidine treatment. Lanes 4 and 5 contain DNA incubated for 36 hr. with 5'-*N*-bromoacetyl oligonucleotide 5, followed by piperidine treatment. (right). Sequence of the oligonucleotide and DNA target site. The major sites of modification are indicated.

Fig. 6. (left) Diagram of linearized plasmid pUCLEU2C indicating location and sequence of triple helical oligonucleotide binding sites. The sites of alkylation and subsequent DNA cleavage are indicated.

(right). Autoradiogram of a 1.0% agarose gel of cleavage products from the reaction of oligonucleotide 5 with ³²P 3'-doubly end-labeled HindIII plasmid pUCLEU2C (6.7 kbp in length). Cleavage products were analyzed on a 1% agarose gel. Lane 1: DNA size markers obtained by digestion with restriction endonucleases. Sizes are indicated to left of band. Lane 2: DNA incubated in the absence of oligonucleotide 5, precipitated, and piperidine treated. Lane 3: DNA incubated with oligonucleotide 5, precipitated, and piperidine treated.

Fig. 7. Scheme for the ligation of alkylation cleavage products from plasmid pUCLEU2C into the complementary Sfi I site of plasmid pUCSfiI. Incubation of plasmid pUCLEU2C with *N*-bromoacetyl oligonucleotide 3, followed by piperidine treatment results in double strand cleavage products with a 3' overhang (CTA-3').

Digestion with the restriction endonuclease EcoR I produces two fragments containing one 5'-AATT and one CTA-3' overhang. Both of these products are complementary with the Sfi I/EcoR I digestion product of pUCSfiI (5'-AATT and TAG-3' overhangs). After circularization and transformation of bacterial cells with the new plasmid, DNA was isolated and sequenced to verify the production of ligatable ends.

Fig. 8. Cleavage of yeast chromosome III by *N*-bromoacetyloligonucleotide 5.

(left). Diagram of yeast chromosome III indicating location and sequence of triple helical oligonucleotide binding sites and the *HIS4* locus used for cleavage analysis. The binding site sequence and sites of alkylation and subsequent DNA cleavage are indicated.

(right). DNA from yeast strains SEY6210 (no target site) and SEY6210C (+ target site) were equilibrated in 1.0 mM $\text{Co}(\text{NH}_3)^{+3}$, 20 mM Hepes pH 7.4, 10 mM EDTA, and 2 μM *N*-bromoacetyloligonucleotide 5. After 24 hr., the plugs were washed with a Tris pH 9.5 (10 mM), EDTA (10 mM) solution three times (10 min. each wash). Plugs were re-equilibrated with 1.0 mM $\text{Co}(\text{NH}_3)^{+3}$, 20 mM Hepes pH 7.4, 10 mM EDTA, and 2 μM *N*-bromoacetyloligonucleotide 5. After a second 24 hr. incubation (37 °C), the plugs were washed in 0.1% piperidine, 100 mM NaCl, 10 mM EDTA, 10 mM Tris pH 8.0, (3 washes 10 min. each) and incubated in 900 μl of this solution for 12 hr. at 55 °C. The plugs were equilibrated with 0.5 x TBE buffer, and loaded into a 1.0% agarose gel for pulsed field gel electrophoresis. The gel was run for 24 hr. at 200 V. Pulsed times were ramped from 10-60 s. for the first 18 hr., and from 60-90 s. over the last 6 hr.

A. Ethidium Bromide stained pulsed field agarose gel. Lane 1: DNA from yeast strain SEY6210 lacking triple helical target site incubated with *N*-bromoacetyloligonucleotide 5. Lane 2: DNA from yeast strain SEY6210C incubated and piperidine treated in the absence of *N*-bromoacetyloligonucleotide 5. Lane 3: DNA from yeast strain SEY6210C incubated with *N*-bromoacetyloligonucleotide 5 and piperidine treated.

B. Autoradiogram of a DNA blot hybridization experiment of gel in A with a 250 bp HIS 4 fragment radiolabeled with ^{32}P by random-primer extension. Locations of the intact chromosome III (340 kbp) and the 110 kbp cleavage fragment are indicated.

