PONTIFICIAE ACADEMIAE SCIENTIARVM SCRIPTA VARIA
70

WORKING GROUP

ON:

MOLECULAR MECHANISMS OF CARCINOGENIC AND ANTITUMOR ACTIVITY

October 21-25, 1986

EDITED BY
CARLOS CHAGAS and BERNARD PULLMAN

ABSTRACT

PONTIFICIA
ACADEMIA
SCIENTIARVM

EX AEDIBVS ACADEMICIS IN CIVITATE VATICANA

MCMLXXXVII

DESIGN OF SYNTHETIC SEQUENCE SPECIFIC DNA BINDING MOLECULES

PETER B. DERVAN
Division of Chemistry and Chemical Engineering
California Institute of Technology
Pasadena, California 91125, USA

Abstract

The design of sequence specific DNA binding molecules has advanced in recent years due, in part, to analytical techniques such as footprinting and affinity cleaving which allow rapid and precise analysis of hundreds of potential DNA binding sites on sequencing gels (Dervan, 1986). *Nona*-N-methylpyrrolecarboxamide, a synthetic analog of the natural product distamycin, binds 11 contiguous base pairs of A,T rich DNA in the minor groove. A synthetic polypeptide, 52 amino acid residues in length and derived from a recombinase protein, binds uniquely 12 base pairs of DNA, most likely in the major and minor groove. The construction of synthetic molecules that bind in the minor and major groove of DNA with incrementally increasing sequence specificity is the first step toward defining a set of rules for the three-dimesional readout of double helical DNA. This may lead to new research tools for use in cancer research, diagnosis of disease states at the level of DNA (oncogenes), and novel chemotherapeutic strategies such as artificial repressors for inactivation of these genes.

1. Introduction

The local structure of right-handed double-helical DNA depends on base sequence (Dickerson *et al.*, 1982; Kennard, 1984; A. H.-J. Wang *et*

al., 1979; Rich *et al.*, 1984; Calladine, 1982; Calladine and Drew, 1984). Many (but not all) low molecular weight natural products isolated as antiviral, antibiotic or anticancer compounds appear to bind in the minor groove of DNA (Gale *et al.*, 1981). Sequence-specific DNA-binding proteins appear to bind, at least in part, in the major groove of DNA (Takeda *et al.*, 1983). Therefore, the design aspects for the construction of sequence-specific DNA-binding molecules focus on understanding the molecular basis for the binding of low molecular weight drugs in the minor groove and of polypeptides derived from proteins in the major groove. The tools of synthetic organic chemistry are used in combination with nucleic acid techniques such as high resolution gel electrophoresis to define, in part, the scope and limitations of this problem (Dervan, 1986).

Examination of natural products that bind to specific sequences in the minor groove of B-DNA reveal that some of these DNA-binding molecules are flat, and this shape allows them to intercalate between the base pairs (Gale *et al.*, 1981). Other natural products with less obvious structural features fit snugly in the minor groove of the DNA helix. We would like to understand the relative contributions of nonbonded stabilizing and destabilizing interactions that allow structurally diverse natural products such as netropsin and distamycin to bind A,T-rich sequences of DNA; and echinomycin, triostin A, bleomycin, actinomycin D, and chromomycin to bind G,C-rich sequences of DNA. A small but significant number of crystal structures of small molecule-oligonucleotide complexes such as actinomycin D (Jain and Sobell, 1972), daunomycin (Quigley *et al.*, 1980), triostin A (Wang *et al.*, 1984), echinomycin (Ughetto *et al.*, 1985), and netropsin (Kopka *et al.*, 1985) are now available. In the absence of crystal structures for other natural products, such as chromomycin bound to duplex DNA, we rely on model building with consensus nucleotide sequences gained from footprinting experiments on several DNA restriction fragments. Plausible models are tested by synthesizing simpler molecules that are believed to contain key recognition features of the more complex natural product. CPK space-filling models of DNA-small molecule complexes are useful for indicating what is unrealistic. As computer graphics model building and adequate molecular mechanics programs become available, there is hope for improvement in this area (Kollman, 1985).

2. The Analytical Problem

There are four bases possible for each nucleotide position on each strand of the DNA helix, and, within the constraints of the A,T and G,C complementary nature of double-helical DNA, for a binding-site size of n base pairs there are $(4^n)/2$ distinguishable sequences for odd n and $(4^n)/2 + (4^{n/2})/2$ for even n. Natural products that are in the molecular weight range of 500 to 2000 are sufficiently large to cover two to six contiguous base pairs. For binding-sites that are two to six base pairs in size, there are 10 to 2080 unique combinations of base pairs or specific binding sites on double helical DNA, respectively (Table 1). During the past few years, a key issue has been the development of the analytical methods to analyze the sequence specificities of either natural or synthetic DNA-binding small molecules. These methods are footprinting (Galas and Schmitz, 1978; Van Dyke *et al.*, 1982; Van Dyke and Dervan, 1982;

Table 1 - *Binding Site Frequency for DNA Binding Molecules.*

Size Site (n)	Unique Sites (N)	Unique Specificity[a]	Unique A·T Sites	A·T Specificity[b]
1	2	5.0×10^{-1}	1	5.0×10^{-1}
2	10	1.0×10^{-1}	3	3.0×10^{-1}
3	32	3.1×10^{-2}	4	1.3×10^{-1}
4	136	7.4×10^{-3}	10	7.4×10^{-2}
5	512	2.0×10^{-3}	16	3.1×10^{-2}
6	2,080	4.8×10^{-4}	36	1.7×10^{-2}
7	8,192	1.2×10^{-4}	64	7.8×10^{-3}
8	32,896	3.0×10^{-5}	136	4.1×10^{-3}
9	131,072	7.6×10^{-6}	256	2.0×10^{-3}
10	524,800	1.9×10^{-6}	528	1.0×10^{-3}
11	2,097,152	4.8×10^{-7}	1,024	4.9×10^{-4}
12	8,390,656	1.2×10^{-7}	2,080	2.5×10^{-4}
13	33,554,432	3.0×10^{-8}	4,096	1.2×10^{-4}
14	134,225,920	7.5×10^{-9}	8,256	6.2×10^{-5}
15	536,870,912	1.9×10^{-9}	16,384	3.1×10^{-5}
16	2,147,516,416	4.7×10^{-10}	32,896	1.5×10^{-5}

([a]) = 1/N; ([b]) = Unique A·T Sites/N.

Scamrov and Beabealashvilli, 1983; Lane *et al.*, 1983; Van Dyke and Dervan, 1983, 1984; Low *et al.*, 1984; Fox and Waring, 1984; Harshman and Dervan, 1985) and affinity cleaving (Schultz *et al.*, 1982; Taylor *et al.*, 1984; Schultz and Dervan, 1983, 1984; Youngquist and Dervan, 1985; Dervan, 1986).

Footprinting and affinity cleaving techniques, which exploit the analytical power of high resolution sequencing gels, allow a rapid and precise solution to the "product analysis" problem and are an underpinning to recent advances in the molecular recognition of DNA area.

A) *Footprinting*

With the availability of restriction endonucleases and techniques that allow the isolation of discrete DNA fragments, uniform DNA substrates are available that have a sufficiently large number of base pairs or combinations of base pairs to be representative of all possible small molecule-binding sites on DNA. With routine enzymatic procedures, DNA fragments (typically 100 base pairs in size or larger) are tagged on one end of one strand (5′ or 3′) with ^{32}P. Footprinting can be carried out with DNA-cleaving agents such as the enzyme DNase I or the synthetic reagent methidiumpropyl-EDTA (MPE), which cleave double-helical DNA at every base position. After cleavage of a ^{32}P-labeled restriction fragment, the set of labeled DNA cleavage fragments differing in length by 1 nucleotide is resolved on a high-resolution denaturing polyacrylamide gel. A bound ligand protects the DNA-binding site from cleavage by covering the base pairs it binds. This is visualized on the autoradiogram of the high-resolution gel as a gap in the ladder of DNA fragments. A chemical sequencing lane run alongside as a marker permits precise identification of these protected regions.

MPE contains the DNA intercalator, methidium, covalently bound by a short hydrocarbon tether to the metal chelator EDTA (Hertzberg and Dervan, 1982, 1984). In the presence of ferrous ion, reducing agents such as dithiothreitol and dioxygen, MPE at micromolar concentrations produces single-strand breaks at 25°C (pH 7.0) in double-helical DNA (Hertzberg and Dervan, 1982, 1984). The synthetic molecule, MPE·Fe(II), is often employed in footprinting experiments because it mimics DNase I as a DNA-cleaving reagent but affords more accurate resolution of binding-site sizes.

The DNA helix accommodates some of these small molecules by changing its shape. How far from the bound site altered DNA structure extends undoubtedly depends on both the DNA binding molecule and the neighboring DNA sequences. DNase I is sensitive to local DNA structure and enzymatic cleavage can be inhibited or enhanced by altered DNA structure. Using DNase I footprinting, several groups have observed enhanced rates of cleavage flanking the binding sites of the antibiotics distamycin, actinomycin, and echinomycin (Scamrov and Beabealashvilli, 1983; Lane *et al.*, 1983; Van Dyke and Dervan, 1983; Low *et al.*, 1984; Fox and Waring, 1984). This enhanced susceptibility to cleavage by DNase I has been interpreted as alteration of the width of the minor groove (Fox and Waring, 1984). DNase I footprinting is proving to be a sensitive technique for determining the extent and sequence dependence of altered DNA structure in solution induced by small molecules proximal and distal to specific binding sites on DNA. Understanding the extent of conformational changes produced by ligand binding, whether they are restricted to the actual binding site or distributed over neighboring regions of DNA, will influence the choice of coupling strategies for synthetic hybrids of different DNA-binding natural products.

B) *Affinity Cleaving*

Attachment of EDTA·Fe(II) to a DNA binding molecule creates an efficient DNA-cleaving molecule (at 25°C and pH 7.0) (Hertzberg and Dervan, 1982, 1984). The EDTA attachment converts a sequence-specific DNA-binding molecule to a sequence-specific DNA-cleaving molecule (Dervan, 1986). Analysis of the cleavage products from a ^{32}P-end-labeled restriction fragment on the autoradiogram of a high-resolution denaturing polyacylamide gel allows the binding locations, site size, and orientation of synthetic molecules on double-helical DNA to be visualized (Dervan, 1986). The resulting cleavage patterns are the positive image visualized on an autoradiogram with respect to the negative image produced by footprinting.

The antibiotic distamycin is a crescent-shaped tripeptide containing three N-methylpyrrolecarboxamides that binds in the minor groove of B-DNA with a strong preference for A,T-rich sequences (Zimmer, 1975; Krey, 1980; Gursky *et al.*, 1982; Zimmer and Walnert, 1986). The EDTA moiety was tethered to the amino or carboxyl terminus of *tris*-N-methylpyrrolecarboxamide to give distamycin-EDTA (DE) and EDTA-distamycin (ED), respectively (Taylor *et al.*, 1984). DE·Fe(II) and ED·Fe(II) at micromolar concentrations cleave pBR322 plasmid DNA at discrete locations

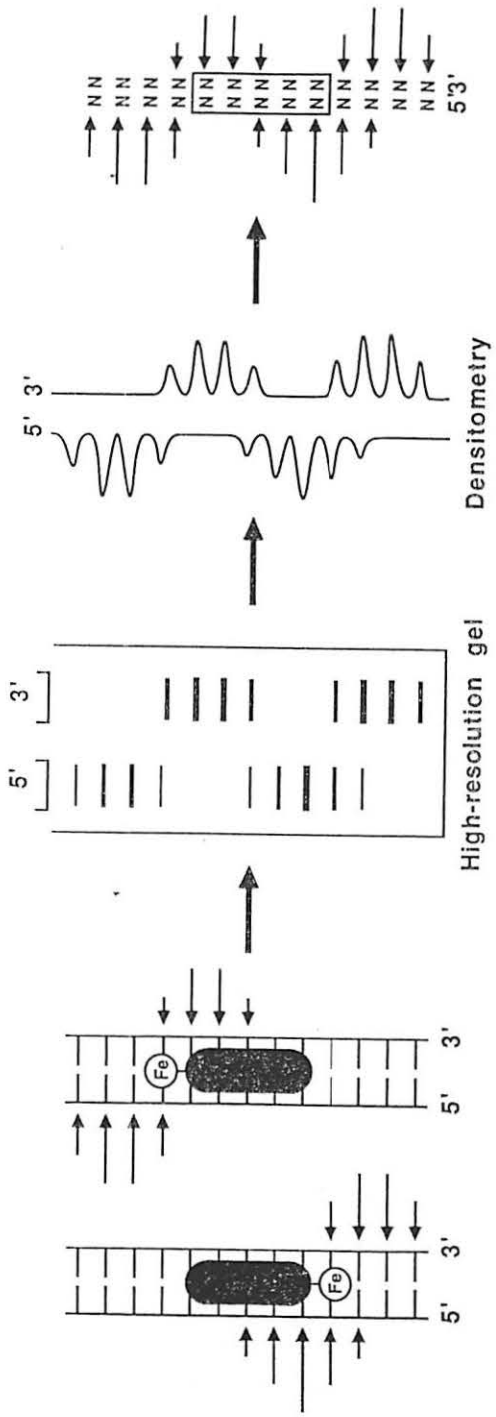

FIG. 1. Scheme for the affinity cleaving method (Dervan, 1986).

in the presence of oxygen and reducing agents such as dithiothreitol. From sequencing gel analyses, it has been found that DE·Fe(II) and ED·Fe(II) afford DNA-cleavage patterns covering four contiguous base pairs adjacent to 5-bp sites consisting of A,T-rich DNA. The multiple contiguous cleavages at each site are taken as evidence for a diffusible oxidizing species, most likely hydroxyl radical (Taylor *et al.*, 1984; Hertzberg and Dervan, 1984). The relative intensity of cleavages on each side of the 5-bp site permits assignment of major and minor orientations of the tripeptide binding unit. The cleavage patterns on opposite DNA strands are asymmetric, shifted to the 3′ side, which can be understood by examination of a model of B-DNA. In the minor groove of right-handed DNA, the proximal deoxyriboses on opposite strands are 2 bp apart to the 3′ side. Assuming that the multiple cleavage events result from a diffusible reactive species, the average position of the EDTA·Fe(II) group is given by the approximate twofold symmetry of the cleavage pattern. From this position, the site of the attached DNA-binding unit can be estimated.

3. Recognition of Large Sequences of A,T-Rich DNA in the Minor Groove

From recent X-ray analysis of the complex of netropsin with the B-DNA dodecamer 5′-CGCGAATTCGCG-3′, Dickerson and co-workers have provided a molecular basis for the sequence-specific recognition of DNA by a *bis*-N-methylpyrrolecarboxamide and, by extension, distamycin (Kopka *et al.*, 1985). Netropsin sits symmetrically in the center of the minor groove of right-handed DNA and displaces the water molecules at the spine of hydration. Each of its three amide groups forms a bridge between adajacent adenine N-3 or thymine O-2 atoms on opposite helix strands. Dickerson and co-workers suggest that the base specificity of netropsin for contiguous A,T-rich sequences in B-DNA is provided not by hydrogen bonding but by close van der Waals contacts between adenine C-2 hydrogens and CH groups on the pyrrole rings of the oligopeptide molecules. Because increased binding-site size would afford increased sequence specificity for DNA-binding molecules, the question arises as to whether higher numbers of N-methylpyrrolecarboxamides in synthetic oligopeptides would fit the natural twist of the B-DNA helix (Arcamone *et al.*, 1969; Zimmer *et al.*, 1983).

Tetra-, penta-, and hexa, hepta-, octa-, and nona-(N-methylpyrrolecarboxamide)s equipped with EDTA (P4E, P5E, P6E, P7E, P8E, and P9E) have been synthesized in our group. The sequence specificities, binding site sizes and orientation preferences have been compared using the affinity cleaving method (Youngquist and Dervan, 1985, 1986). The homologous oligopeptide-EDTA·Fe(II) molecules cleave restriction fragments at common locations rich in A,T that differ incrementally in the size of the binding site (Youngquist and Dervan, 1985, 1986). From analysis of the cleavage patterns visualized by high-resolution denaturing gel electrophoresis, the oligopeptides with four-nine N-methylpyrrolecarboxamide units and containing five-ten amide groups bind sites of A,T-rich DNA consisting of six-eleven contiguous base pairs, respectively. The general rule of n amides affording binding-site sizes of $n+1$ base pairs is explained by the solid state structure of the netropsin: DNA duplex.

From the relative intensities of the cleavage patterns flanking the

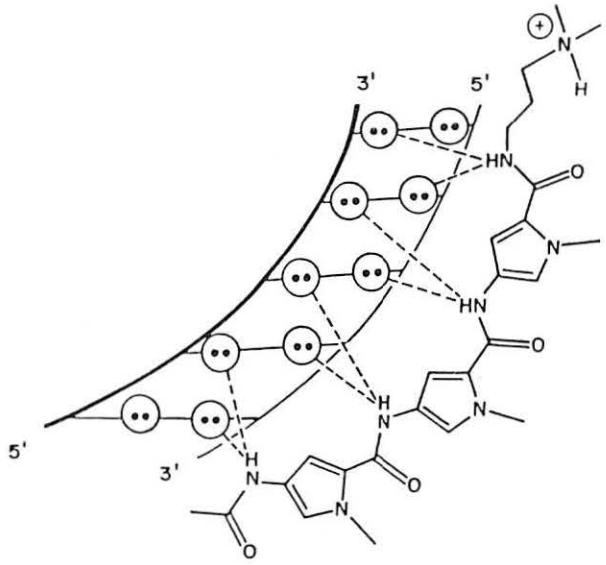

FIG. 2. Model for *tris*-N-methylpyrrolecarboxamide binding in the minor groove of DNA at A,T rich sequences 5 bp in size. (See Kopka *et al.*, 1985) Circles with two dots represent lone pairs of electrons on N3 of adenine and 02 of thymine. Dotted lines are hydrogen bonds.

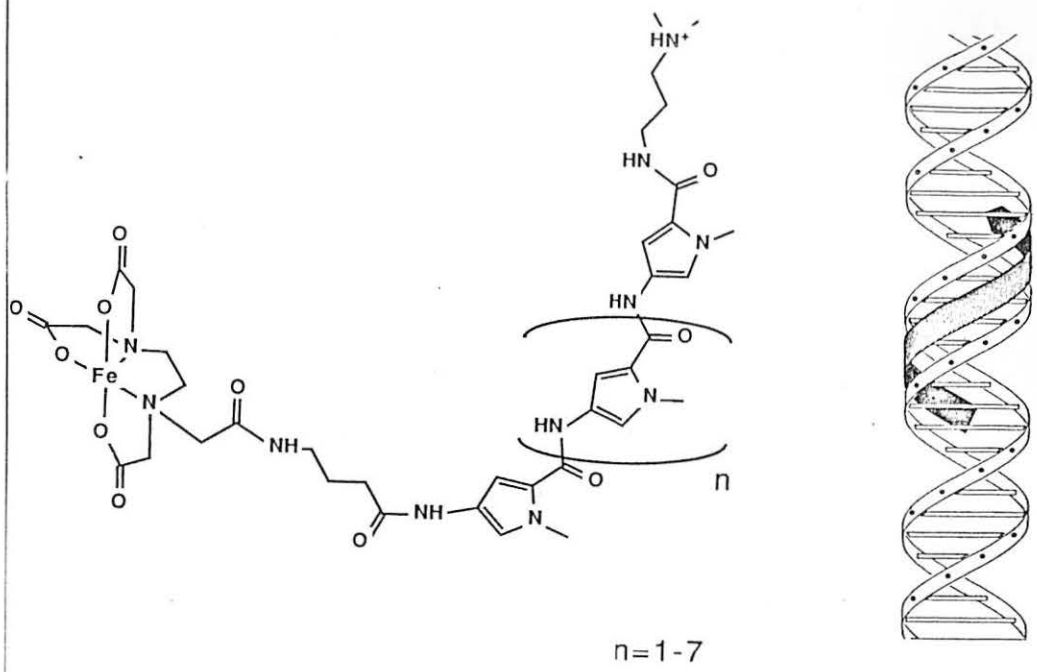

FIG. 3. *Oligo*-N-methylpyrrolecarboxamides equipped with iron-EDTA at the amino end (Youngquist and Dervan, 1985, 1986).

binding site, the orientation preference of the oligopeptide at each binding site can be estimated as a function of local sequence, flanking sequences, and number of N-methylpyrrolecarboxamide units. Dickerson has shown that, although netropsin binding neither unwinds nor elongates the dodecamer, it does force open the minor groove by 0.5 to 2.0 Å and bends back the helix axis by 8° across the region of attachment. One explanation for nonequivalent binding orientation on an AT-rich binding site that lacks twofold symmetry is that the narrowness of the minor groove in B-DNA differs with local DNA sequence. For example, DE has an orientation preference for the amino end of the tripeptide to the 5′ side of the sequence 5′-agAAATTgc-3′. DE has no orientation preference for the sequence 5′-ttAAATTgc-3′ (Taylor *et al.*, 1984; Youngquist and Dervan, 1985). Since the binding sites are the same but occur in different locations

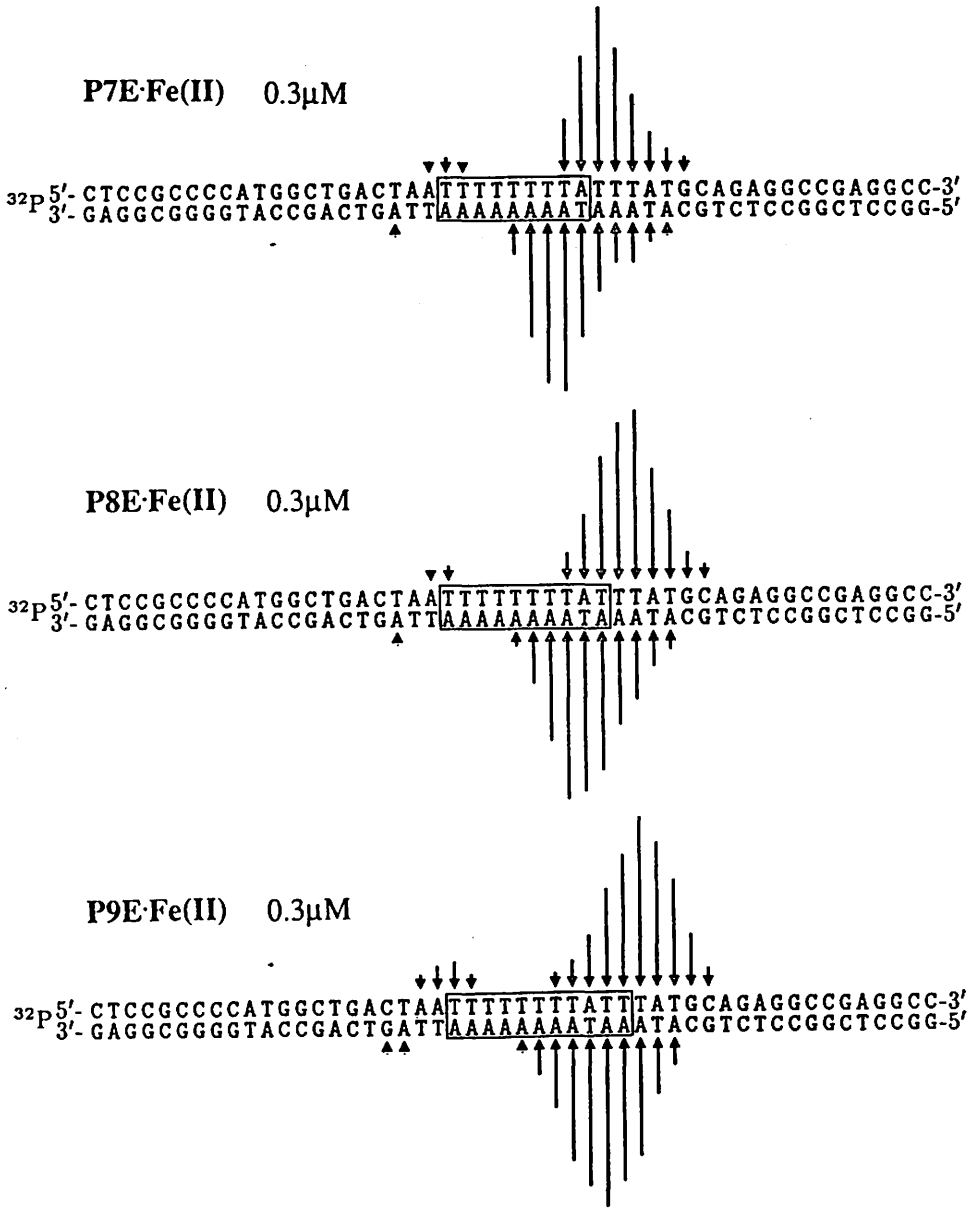

FIG. 4. Histograms of the DNA cleavage patterns on a DNA restriction fragment by P7E·Fe(II), P8E·Fe(II) and P9E·Fe(II). (Youngquist and Dervan, 1986). Boxes define binding site location and size based on model described in Taylor, *et al.*, 1984; Youngquist and Dervan, 1985; Dervan, 1986.

on DNA, we conclude that different local structures can be conferred on identical binding sites by flanking sequences.

Netropsin and distamycin have been characterized as molecules that bind preferably to A,T-rich regions of DNA. It is known from equilibrium binding studies of distamycin analogs that homopolymer dA·dT sequences are preferred over alternating d(A-T)·d(A-T) copolymer sequences. From footprinting and affinity cleaving data we find that G,C base pairs are sometimes permissible in the preferred binding sites of netropsin, distamycin, and higher oligo(N-methylpyrrolecarboxamide) homologs on DNA. (Youngquist and Dervan, 1985) Perhaps the close van der Waals nonbonded contacts between the pyrrole CH and the -NH_2 group of guanine are adjustable or are not identical for every pyrrole position on the bound crescent-shaped oligopeptide. Moreover, there are several pure A,T sites that are not strong binding sites for the oligo(N-methylpyrrolecarboxamide)s.

According to the $n+1$ rule, the minimum recognition unit for the N-methylpyrrolecarboxamide on B-DNA is 2 bp. If the recognition elements for the carboxamide NH are on adjacent residues on opposite helix strands, there are ten bridged base possibilities: AA, AT, AC, AG, TT, TC, TG, CC, GG, and CG. The data from affinity cleaving studies reveal that the preferences of bridge hydrogen bonds between adjacent bases on opposite helix strands decrease in the following order: AT≫AA>TT>AC,TC,TG≫AG,CG,CC,GG.

4. Coupled DNA Binding Units of Similar Specificity in the Minor Groove

Undoubtedly there is an upper limit where oligo(N-methylpyrrolecarboxamide)s will no longer fit the natural twist of the B helix. For sequence-specific DNA-binding molecules that read large sequences of double-helical DNA, there will be a need to *couple* DNA-binding units derived from natural products of similar (or mixed) base-pair specificities. For the success of this coupled DNA-binding unit approach, the base-specific recognition elements of each subunit and the linkers connecting them must be compatible with the same groove and conformational state of the DNA.

The initial design attempts for large sequences of A,T-rich DNA involved the construction of dimers of di- and tripeptides connected by flexible hydrocarbons tethers (Khorlin *et al.*, 1980; Schultz and Dervan, 1983). From affinity cleaving experiments, a tripeptide dimer with the

C_7 linker, bis(EDTA-distamycin)Fe(II) [BED·Fe(II)], binds a 9-bp A,T-rich site (5′-ATTTTTATA-3′), a result consistent with simultaneous binding. However, this dimer also binds a 5-bp site (5′-AATAA-3′), suggesting that the C_7 hydrocarbon tether allows both dimeric and monomeric binding modes (Schultz and Dervan, 1983).

A shorter linker, such as the diamide of fumaric acid, more closely mimics the N-methylpyrrolecarboxamide DNA-binding unit with regard to shape and curvature between the amide NH's. Bis(EDTA-distamycin) fumaramide (BEDF), which is a crescent-shaped octamide containing two N-methylpyrrole tripeptide units coupled at the amino termini via a C_4 tether, fumaric acid, reveals major cleavage sites flanking two AT-rich sequences, 5′-ATTTTTATA-3′ and 5′-ATAATAAT-3′ (Youngquist and Dervan, 1985). The observation of 8- to 9-bp binding in the absence of 5-bp binding for BEDF suggests that the tripeptides are binding exclusively simultaneously on double-helical DNA.

5. Coupled DNA Binding Units of Mixed Specificity in the Minor Groove

Several natural products that bind double-helical DNA have the structural-binding feature of being both an intercalator and groove binder. The natural product actinomycin D consists of an aromatic chromophore, phenoxazone, coupled to two identical cyclic pentapeptide lactones (Jain and Sobell, 1972; Takusagawa *et al.*, 1982). Actinomycin is an intercalator-groove binder that unwinds DNA by 26°. The cyclic pentapeptides fit snugly above and below the intercalating ring in the minor groove of DNA. Actinomycin binds 4 bp with a preference for 5′NGCN-3′ sequences (Van Dyke *et al.*, 1982; Van Dyke and Dervan, 1982; Scamrov and Beabealashvilli, 1983; Lane *et al.*, 1983).

We tested whether the intercalator portion of actinomycin, phenoxazone, can be covalently linked to a different minor groove binder, distamycin, in such a way that the sequence specificity of *both* moieties is retained. The two cyclic pentapeptides of actinomycin were replaced with the tripeptide from distamycin and the postulated key carbonyl amide guanine NH recognition element of actinomycin was retained (Dervan and Sluka, 1986). As judged from distance and steric considerations from model building studies, the distamycin tripeptide was connected to the phenoxazone by a glycine tether (Dervan and Sluka, 1986). The resulting bis(distamycin)phenoxazone is a synthetic hybrid groove binder-intercalator

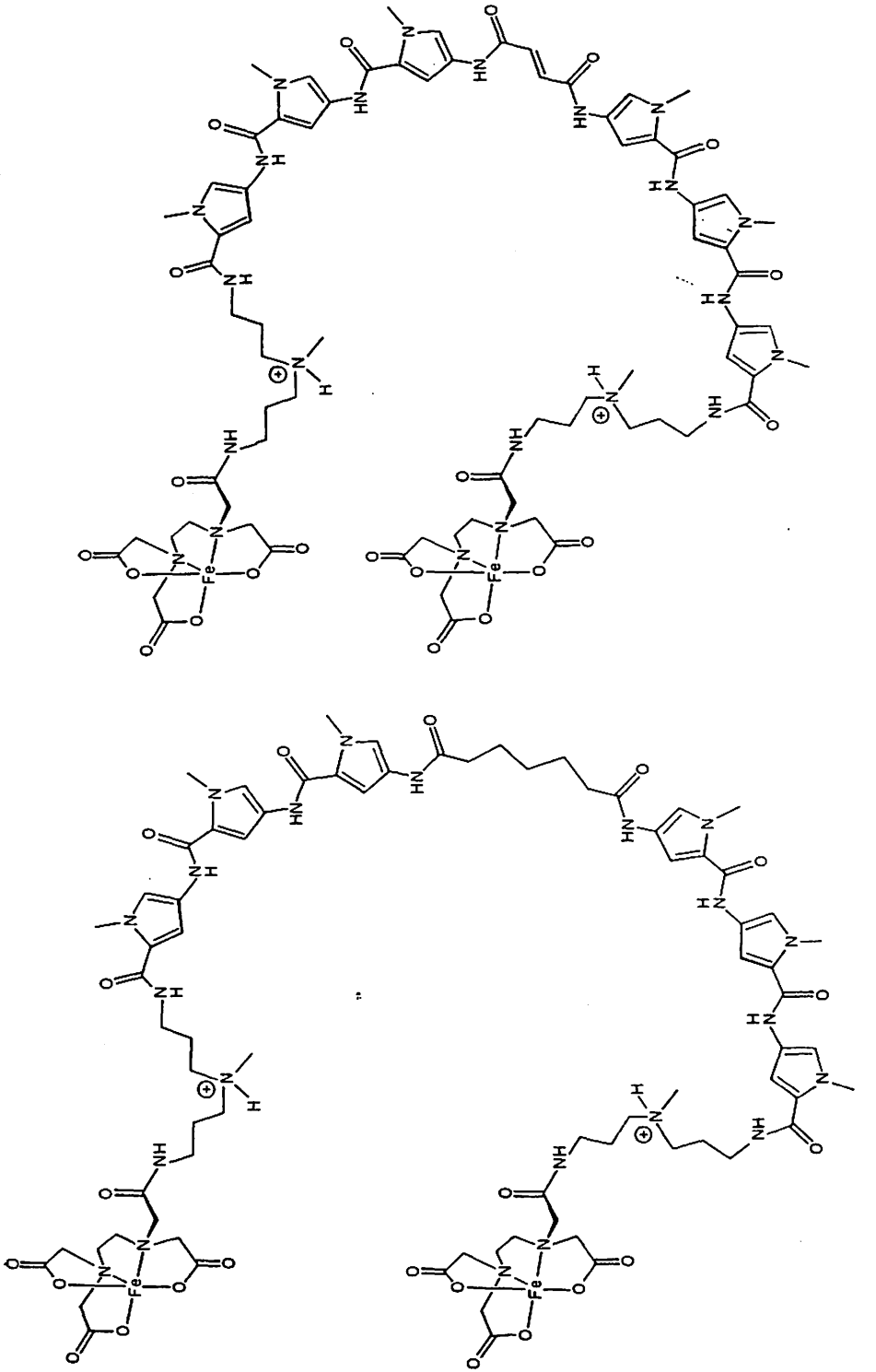

Fig. 5. Dimers of distamycin connected by C7 and C4 linkers (Schultz and Dervan 1983, Youngquist and Dervan, 1985).

that in a formal sense might bind 10 bp of DNA having the sequence $(A,T)_4(G,C)_2(A,T)_4$. The sequence specificity of bis(distamycin)phenoxazone on DNA restriction fragments was tested by the affinity cleaving method (Dervan and Sluka, 1986).

The DNA cleavage patterns reveal a major cleavage site flanking the 10-bp sequence, 5′-TATAGGTTAA-3′ (Dervan and Sluka, 1986). One interpretation of the data is that the tripeptides are binding simultaneously at A,T-rich sequences 4 bp in size flanking the central 5′GG-3′ phenoxazone binding site. Although not a proof, this would be consistent with the groove binder-intercalator-groove binder mode. However, at two other sites, single-cleavage loci were observed consistent with one tripeptide or possibly a tripeptide-phenoxazone binding at these sites (Dervan and Sluka, 1986). Perhaps at these sites there is intercalation of the phenoxazone with sequence dependent and unequal local distortion on both sides of the intercalation site which would make one site incompatible with the tripeptide groove binder. Because there are 524,800 unique sequences of doublehelical DNA that are 10 bp in size, it is likely that the optimal 10-bp recognition site for bis(distamycin)phenoxazone has not yet been identified.

This limited success with a DNA-binding molecule that is a hybrid of two natural products makes it appear possible that groove binders can be mixed and matched with intercalators and that A,T words can be coupled with G,C words. The choice of a linker connecting the recognition elements may be a critical design feature with regard to simultaneous binding of all moieties. Perhaps further efforts should focus on the synthesis of "mixed specificity" molecules that are *exclusively* groove binders. This might minimize distortion on the DNA partly responsible for the incompatibility of coupled DNA binding units such as intercalators and groove binders. Because our understanding of G,C recognition in the minor groove is not as well advanced as A,T recognition, a challenge for the future with regard to minor groove recognition is the design of G,C minor groove binding moieties.

6. Alteration of the Specificity of Distamycin from Pure (A,T) to Mixed (G,C/A,T)

Recently, we found that replacement of one N-methylpyrrolecarboxamide unit in the tripeptide portion from distamycin with pyridine-2-carboxamide affords a new synthetic DNA minor groove binding molecule,

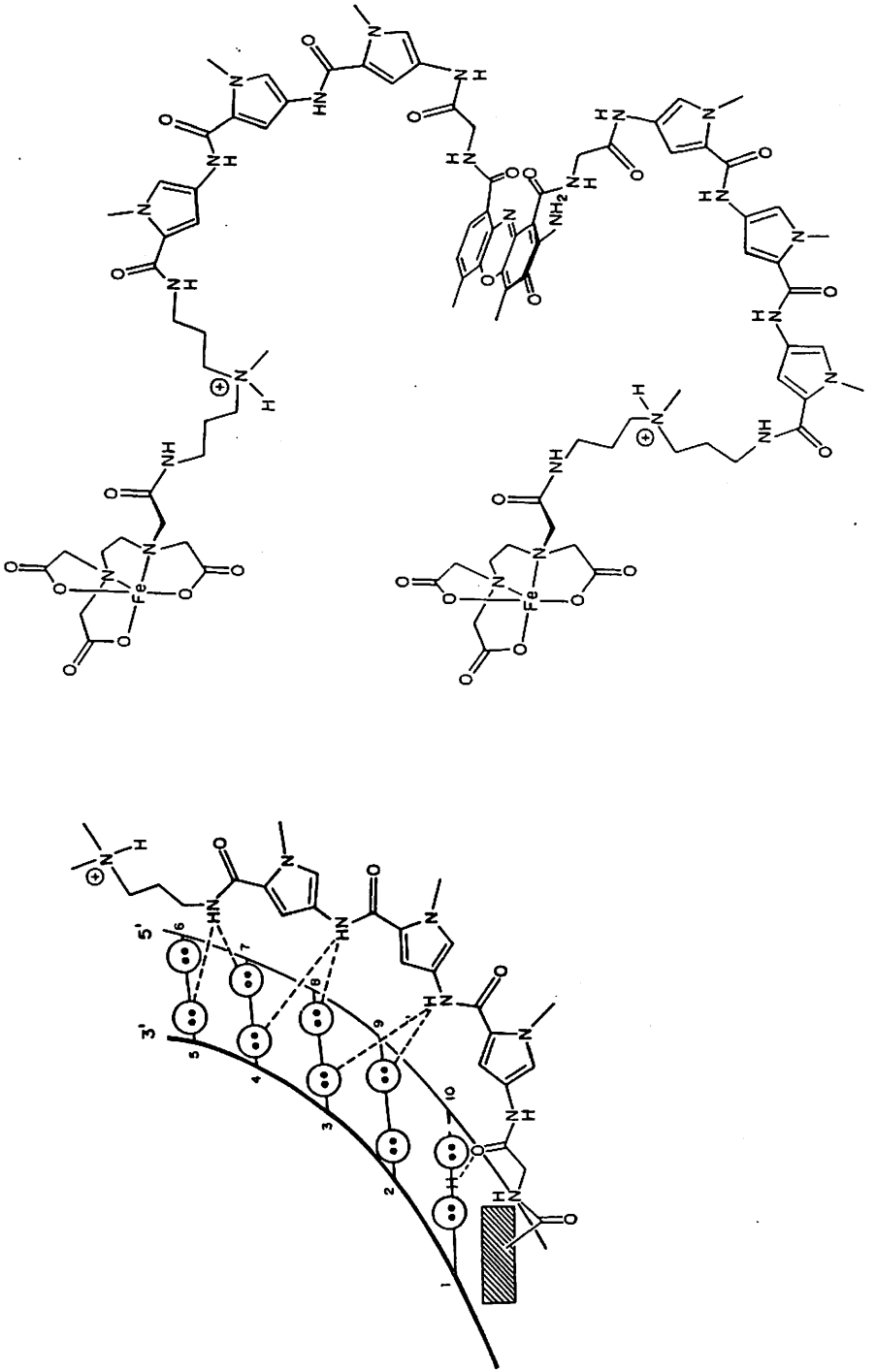

Fig. 6. (*left*) Model for a hybrid intercalator-groove binder, distamycinphenoxazone. Circles with two dots represent lone pairs of electrons on N3 of adenine and 02 of thymine. Circles with two dots and H interposed represent lone pairs of electrons on N3 of guanine and 02 of cytosine flanking the NH_2 of guanine at the edges of the base pairs on the floor of the minor groove of righ-handed DNA. Dotted lines are hydrogen bonds. (*right*) Bis [Fe(II)·EDTA-distamycin]phenoxazone BEDP·Fe(II) (Dervan and Sluka, 1986).

pyridine-2-carboxamidenetropsin (2PyN), that now accepts G,C/A,T base pairs *in preference* to pure A,T stretches of DNA (Wade and Dervan, 1986). Pyridine-3-(or pyridine-4)-carboxamidenetropsin (3-PyN and 4-PyN) do *not* show this effect supporting a model where the lone pair of electrons on the pyridine nitrogen is responsible, at least in part, for the G,C recognition (Wade and Dervan, 1986).

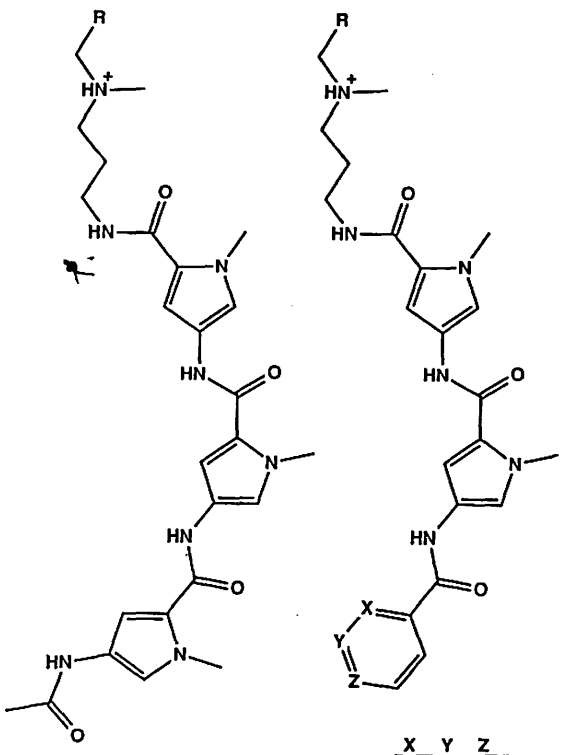

FIG. 7. (*left*) Analog of distamycin. (*right*) Synthetic groove binders, pyridinecarboxamidenetropsin.

7. POLYPEPTIDE RECOGNITION IN THE MAJOR AND MINOR GROOVE OF DNA

In recent years x-ray crystal structures of a number of sequence specific DNA binding repressor proteins have become available (Ohlendorf *et al.*, 1982; Takeda *et al.*, 1983; Pabo and Lewis, 1982; Pabo and Sauer, 1984; Lewis *et al.*, 1983; Steitz *et al.*, 1983; McKoy *et al.*, 1982; McKoy and Steitz, 1981). Comparison of the three-dimensional structures has led to the postulate that a conserved helix-turn-helix motif is intimately involved in sequence-specific DNA recognition. The "solvent exposed" face of the second helix is believed to lie on the floor of the major groove of B-DNA and make sequence-specific contacts with the exposed edges of the bases. Comparison of the amino acid sequence of other DNA binding proteins to the postulated binding domains of repressor proteins suggests that the helix-turn-helix motif interacting with the major groove of B-DNA may be a common structural feature employed by some DNA binding proteins.

To apply the affinity cleaving method to the protein:DNA recognition area, we have developed a general method to attach EDTA to a specific amino acid residue of a synthetic polypeptide fragment. Because the polypeptide-EDTA·Fe(II) should be oriented in one direction on the DNA helix, the affinity cleaving data would reveal the binding location and the orientation of the peptide on the DNA. Importantly, the asymmetry of the cleavage pattern would also reveal the location of that particular amino acid residue equipped with EDTA·Fe(II). For cleavage in the minor groove the pattern is asymmetric to the 3′ side. For cleavage in the major groove the pattern is asymmetric to the 5′ side.

We have synthesized a polypeptide, 52 amino acid residues in length, derived from the DNA binding domain of the Hin site-specific recombinase protein, that is equipped with EDTA at the N-terminal end (Sluka *et al.*, 1986). The experimental work will not be documented here, but early indications from affinity cleaving experiments reveal that the synthetic polypeptide-EDTA with two structural and functional domains (DNA recognition/DNA cleavage) binds 12 base pairs of DNA, most likely both in the major and minor groove (Sluka *et al.*, 1986).

8. SUMMARY

The design of synthetic sequence-specific binding molecules for double helical DNA has advanced in recent years due in part to the development

of analytical techniques such as footprinting and affinity cleaving for analyzing rapidly and precisely on sequencing gels hundreds of potential DNA binding sites. In the minor groove, the synthetic *nona*-N-methyl-pyrrolecarboxamide binds 11 contiguous base pairs of A,T DNA. This is comparable to the specificity of a restriction enzyme which recognizes uniquely 6 bp of DNA. Corresponding success in the major groove with synthetic polypeptides puts unique specificity at the 12 base pair recognition level within reach. This is an encouraging first step toward defining a set of rules for the three-dimensional read-out of double helical DNA. A long range goal is *unique* specificity at the 15 base pair level which would make possible the design of artificial repressors for the inactivation of oncogenes.

ACKNOWLEDGEMENTS

We are very grateful for the generous support of the National Institutes of Health, the American Cancer Society, the National Foundation for Cancer Research, the Burroughs-Wellcome Company, and the Smith Kline Beckman Corporation.

REFERENCES

ARCAMONE F., NICOLELLA V., PENCO S. and REDAELLI S., « Gazz. Chim. Ital. », *99*, 632 (1969).

CALLADINE C.R., « J. Mol. Biol. », *161*, 343 (1982).

CALLADINE C.R. and DREW H.R., ibid., *178*, 773 (1984).

DERVAN P.B., « Science », *232*, 464 (1986).

DERVAN P.B. and SLUKA J., In: *Internat. Kyoto Conf. on Organic Chem. Proceedings*, « New Synthetic Methodology and Functionally Interesting Compounds », (Elsevier, Amsterdam), pp. 307-322 (1986).

DICKERSON R.E., DREW H.R., CONNER B.N., WING R.M., FRATINI A.V. and KOPKA M.L., « Science », *126*, 475 (1982).

DICKERSON R.E., « J. Mol. Biol. », *166*, 419 (1983).

FOX K.R. and WARING M.J., « Nucleic Acids Res. », *12*, 9271 (1984).

GALAS D.J. and SCHMITZ A., « Nucleic Acids. Res. », *5*, 3157 (1978).

GALE E.F., CUNDLIFFE C., REYNOLDS P.-E., RICHMOND M.H. and WARING M.J., *The Molecular Basis of Antibiotic Action* (Wiley, New York), pp. 258-401 (1981).

GURSKY G.V., ZASEDATELEV A.S., ZHUZE A.L., KHORLIN A.A., GROKHOVSKY S.L., STETTSOV S.A., SUROVAYA A.N., NIKITEN S.M., KRYLOV A.S., RETCHINSKY V.O., MIKHALILOV M.V., BEABEALASHVILLI R.S. and GOTTIKK B.P., « Cold Spring Harbor Symp. Quant. Biol. », *74*, 367 (1982).

HARSHMAN K. and DERVAN P.B., « Nucleic Acids Res. », *13*, 4825 (1985).

HERTZBERG R.P. and DERVAN P.B., « J. Am Chem. Soc. », *104*, 313 (1982).

HERTZBERG R.P. and DERVAN P.B., « Biochemistry », *23*, 3934 (1984).

JAIN S.C. and SOBELL H.M., « J. Mol. Biol. », *296*, 1 (1972).

KENNARD O., « Pure Appl. Chem. », *56*, 989 (1984).

KHORLIN A.A. *et al.*, « FEBS Lett. », *118*, 311 (1980).

KOLLMAN P., « Acc. Chem. Res. », *18*, 105 (1985).

KOPKA M.L., YOON C., GOODSELL D., PJURA P. and DICKERSON R.E., « Proc. Natl. Acad. Sci. U.S.A. », *82*, 1376 (1985).

KREY A.K., In: *Progress in molecular and Subcellular Biology*, F.N. Nahn, Ed. (Springer-Verlag, New York), vol. *7* (1980).

LANE M., DABROWIAK J.C., VOURNAKIS J.N., « Proc. Natl. Acad. Sci. U.S.A. », *80*, 3260 (1983).

LOW C.M.L., DREW H.R. and WARING M.J., « Nucleic Acids Res. », *12*, 4865 (1984).

OHLENDORF D.H., ANDERSON W.F., FISHER R.G., TAKEDA Y. and MATTHEWS B.W., « Nature », *298*, 718 (1982).

PABO C.O. and LEWIS M., « Nature », 443-447 (1982).

PABO C.O. and SAUER R.T., « Ann. Rev. Biochem. », *53*, 293 (1984).

QUIGLEY G., WANG A., UGHETTO G., VAN BROOM J., RICH A., « Proc. Natl. Acad. Sci. U.S.A. », *77*, 7204 (1980).

RICH A., NORDHEIM A. and WANG A. H.-J., « Ann. Rev. Biochem. », *53*, 791 (1984).

SCAMROV A.V. and BEABEALASHVILLI R. Sh., « FEBS Lett. », *164*, 97 (1983).

SCHULTZ P.G., TAYLOR J.S., DERVAN P.B., « J. Am. Chem. Soc. », *104*, 6861 (1982).

SCHULTZ P.G. and DERVAN P.B., « J. Am. Chem. Soc. », *105*, 7748 (1983).

SCHULTZ P.G. and DERVAN P.B., « J. Biomol. Struc. Dynam. », *1*, 1133 (1984).

SCHULTZ P.G. and DERVAN P.B., « Proc. Natl. Acad. Sci. U.S.A. », *80*, 6834 (1984).

SLUKA J., HORVATH S., BRUIST M., SIMON M., DERVAN P.B., unpublished observations, 1986.

TAKEDA Y., OHLENDORF D.H., ANDERSON W.F. and MATTHEWS B.W., « Science », *221*, 1020 (1983).

TAKUSAGAWA F., DABROW M., NEIDLE S. and BERMAN H.M., « Nature (London) », *296*, 466 (1982).

TAYLOR J.S., SCHULTZ P.G. and DERVAN P.B., « Tetrahedron », *40*, 457 (1984).

UGHETTO G., WANG A. H.-J., QUIGLEY G.J., VAN DER MAREL G.A., VAN BOOM J.H. and RICH A., « Nucleic Acids Res. », *13*, 2305 (1985).

VAN DYKE M.W., HERTZBERG R.P. and DERVAN P.B., « Proc. Natl. Acad. Sci U.S.A. », *79*, 5470 (1982).

VAN DYKE M.W. and DERVAN P.B., « Cold Spring Harbor Symp. Quant. Biol. », *47*, 347 (1982).

VAN DYKE M.W. and DERVAN P.B., « Biochemistry », *22*, 2373 (1983).

VAN DYKE M.W. and DERVAN P.B., « Nucleic Acids Res. », *11*, 5555 (1983).

VAN DYKE M.W. and DERVAN P.B., « Science », *225*, 1122 (1984).

WADE W. and DERVAN P.B., unpublished observations, 1986.

WANG A. H.-J., QUIGLEY G.J., KOLPAK F.J., CRAWFORD J.H., VAN BOOM J.H., VAN DER MAREL G.A. and RICH A., « Nature (London) », *282*, 680 (1979).

WANG A. H.-J., GIOVANNI U., QUIGLEY G.J., HAKOSHIMA T., VAN DER MAREL G.A., VAN BOOM J.H. and RICH A., « Science », *225*, 1115 (*1984).

WARING M., « J. Mol. Biol. », *54*, 247 (1970).

YOUNGQUIST R.S. and DERVAN P.B., « J. Am. Chem. Soc. », *107*, 5528 (1985).

YOUNGQUIST R.S. and DERVAN P.B., « Proc. Natl. Acad. Sci. U.S.A. », *82*, 2565 (1985).

YOUNGQUIST R.S. and DERVAN P.B., unpublished work, 1986.

ZIMMER C., In: *Progress in Nucleic Acids Research and Molecular Biology*, N.E. Cohn Ed. (Academic Press, New York), pp. 285-318 (1975).

ZIMMER C., LUCK G., BIRCH-HIRSCHFELD E., WEISS R., ARCAMONE F. and GUSCHLBAUER W., « Biochem. Biophys Acta », *741*, 15 (1983).

ZIMMER C. and WALNERT V., « Prog. Biophys Molec. Biol. », *47*, 31 (1986).