Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 2565–2569, May 1985
Biochemistry

Sequence-specific recognition of B-DNA by oligo(*N*-methylpyrrolecarboxamide)s

[netropsin/distamycin/oligopeptide-EDTA·Fe(II)/affinity cleaving]

R. SCOTT YOUNGQUIST AND PETER B. DERVAN*

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

Communicated by J. L. Beauchamp, December 10, 1984

ABSTRACT **Four homologous oligopeptide-EDTA molecules, tri-, tetra, penta-, and hexa(*N*-methylpyrrolecarboxamide)-EDTA, in the presence of Fe(II), O_2, and dithiothreitol, cleave ^{32}P-end-labeled restriction fragments from plasmid pBR322 DNA at common locations rich in A·T base pairs that differ in the size of the binding site. From analysis of the cleavage patterns visualized by high-resolution denaturing gel electrophoresis, the oligopeptides with three, four, five, and six *N*-methylpyrrolecarboxamide units, containing four, five, six, and seven amide NHs, bind sites of A·T-rich DNA consisting of five, six, seven, and eight contiguous base pairs, respectively. The general rule of *n* amides affording binding site sizes of *n* + 1 base pairs is consistent with the oligopeptides binding in the minor groove of right-handed DNA, with the amide NH groups forming bridges between the adjacent N-3 and O-2 atoms of adenine or thymine on opposite strands of the DNA helix.**

Netropsin and distamycin A are antibiotics active against fungi, bacteria, and viruses. These di- and tripeptides show a marked preference for DNA rich in adenine (A) and thymine (T) and have binding sites four to six base pairs in size (1–11). From a recent x-ray analysis of the complex of netropsin with the B-DNA dodecamer of sequence C-G-C-G-A-A-T-T-C-G-C-G, Dickerson and coworkers provide a molecular basis for the sequence-specific recognition of DNA by netropsin, and by extension, distamycin (12). They find that netropsin sits symmetrically in the center of the minor groove of right-handed DNA and displaces the water molecules of the spine of hydration (12). Each of its three amide NH groups forms a bridge between adjacent adenine N-3 or thymine O-2 atoms on opposite helix strands (12). This explains recent solution data that distamycin analogues having *n* amides characteristically bind to *n* + 1 successive base pairs (13–15). Dickerson and coworkers suggest that the base specificity of netropsin for contiguous sequences of A·T base pairs 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 (12). Because *N*-methylpyrrolecarboxamides could be used as recognition elements for the design of synthetic sequence-specific DNA-binding molecules, the question arises whether higher numbers of contiguous *N*-methylpyrrolecarboxamides in synthetic oligopeptides would still fit the natural twist of the B-DNA helix (13–18).

Affinity Cleaving. Attachment of DNA-cleaving moieties, such as EDTA·Fe(II) (19, 20), to DNA-binding molecules followed by DNA cleavage pattern analysis using high-resolution denaturing gel electrophoresis is a useful method for determining the binding locations, site size, and orientation of small molecules on native DNA (14–18). The resulting cleavage patterns are simply the positive image visualized on an autoradiogram with respect to the negative image produced by cleavage inhibition patterns ("footprinting") (17) (Fig. 1). We have synthesized for comparison the tri-, tetra-, penta- and hexa(*N*-methylpyrrolecarboxamide)s equipped with EDTA (Fig. 2). From the cleavage patterns on two ^{32}P-end-labeled restriction fragments from plasmid pBR322 DNA, we have analyzed their binding sites, which are composed of A·T-rich DNA. We find that the binding sites for the tri-, tetra-, penta- and hexapeptides are five, six, seven, and eight base pairs, respectively.

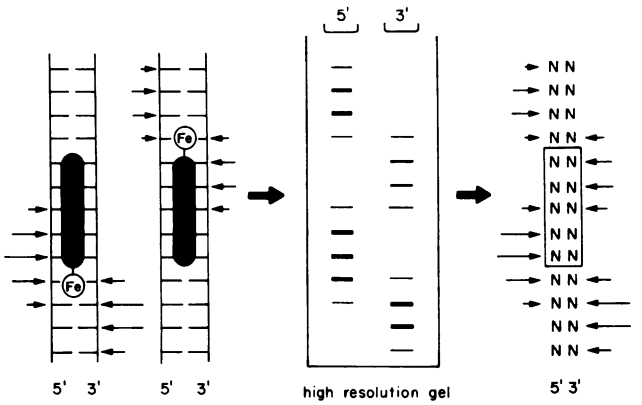

FIG. 1. Scheme for the affinity cleaving method. (*Left*) Oligopeptide-EDTA·Fe(II) can assume two orientations, affording asymmetric cleavage patterns on opposite strands of a 5′- (and 3′)-^{32}P-end-labeled restriction fragment. Lengths of arrows indicate frequency of cleavage. (*Center*) Cleavage patterns on a high-resolution denaturing gel. (*Right*) Assignment of binding site boundaries based on the model in Fig. 5

MATERIALS AND METHODS

The general synthetic methods used for the construction of the oligo(*N*-methylpyrrolecarboxamide)s equipped with EDTA have been described for the case of DE (15). The full synthetic details will be published elsewhere. The DNA used in this investigation was bacterial plasmid pBR322, whose entire sequence is known (21, 22). The ^{32}P-end-labeled restriction fragments were prepared as previously described (17). Calf thymus DNA (Sigma) was sonicated, deproteinized, and extensively dialyzed.

High-Resolution Denaturing Gel. Each oligopeptide-EDTA·Fe(II) molecule was prepared by mixing DE, P4E, P5E, or P6E at 10 mM with one equivalent of $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ and diluting immediately. The cleavage reactions were run with >600 cpm of ^{32}P-end-labeled restriction fragments made up to a total final DNA concentration of 100

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: DE, tri(*N*-methylpyrrolecarboxamide)-EDTA (distamycin-EDTA); P4E, P5E, and P6E, tetra-, penta-, and hexa(*N*-methylpyrrolecarboxamide)-EDTA; bp, base pair(s).

*To whom reprint requests should be addressed.

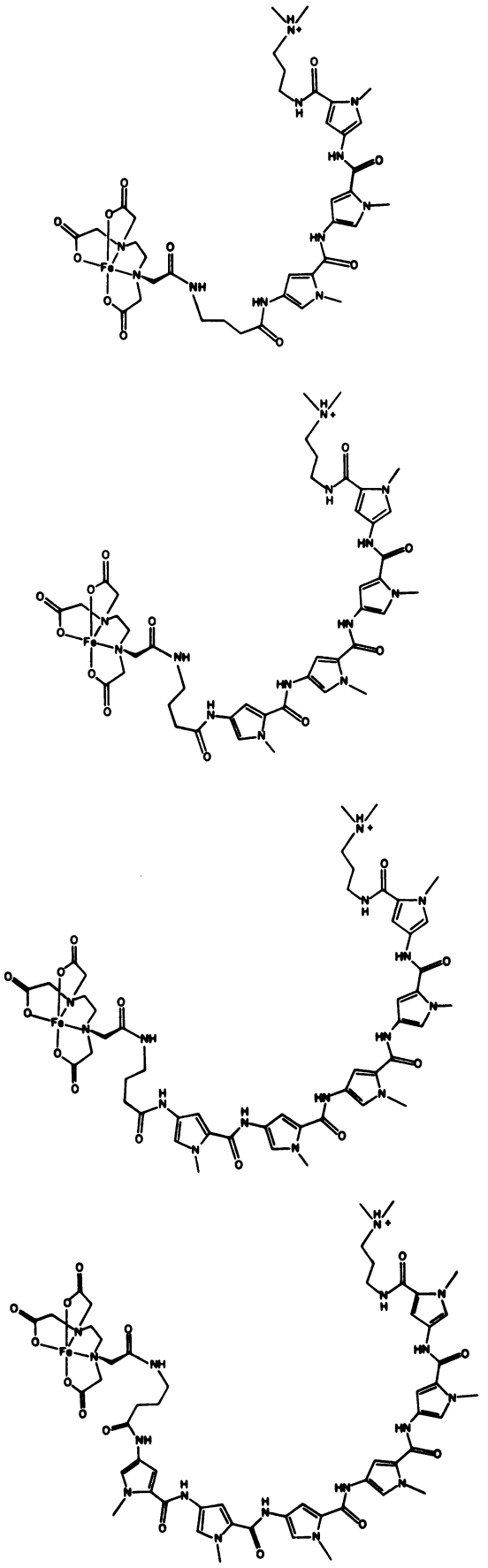

FIG. 2. Tri-, tetra-, penta-, and hexa(*N*-methylpyrrolecarboxamide)-EDTA·Fe(II). Without iron, the reagents are abbreviated DE, P4E, P5E, and P6E, respectively, DE indicating distamycin-EDTA.

μM (in base pairs) with sonicated calf thymus DNA. To 9 μl of DNA/buffer solution, 3 μl of oligopeptide-EDTA·Fe(II) was added. Each oligopeptide-EDTA·Fe(II) complex was allowed to equilibrate with the buffered DNA solution. Cleavage was initiated by the addition of 3 μl of 25 mM dithiothreitol to the reaction mixture to give a total reaction volume of 15 μl. The reaction buffer was prepared to final concentrations of 40 mM Tris base and 5 mM sodium acetate and was adjusted to pH 7.9 with acetic acid. Final concentrations are given in the legend of Fig. 3. The reactions were run for 1 hr. The reaction mixtures were frozen, lyophilized, and suspended in 4 μl of 100 mM Tris borate/formamide 50% (vol/vol) at pH 8.3. The samples were heated/denatured and loaded onto a 0.4-mm-thick, 40-cm-long, 8% polyacrylamide (1:20 crosslinked)/50% urea gel and electrophoresed at 1000 V. The gels were autoradiographed at −50°C and the cleavage patterns were quantified by densitometry.

RESULTS

DNA Cleavage Pattern Analysis. The cleavage sites for each of the oligopeptide-EDTA·Fe(II) molecules on ^{32}P-end-labeled DNA restriction fragments were resolved by using high-resolution gel electrophoresis (Fig. 3). The location and size of the binding site for each oligopeptide could be visualized from autoradiograms of the gels, which were analyzed by densitometry (Fig. 4). Cleavage loci are produced flanking A·T-rich sites. These multiple cleavage patterns, consisting of four or five contiguous base pairs, are asymmetric, shifted to the 3′ side on each DNA strand. The cleavage pat-

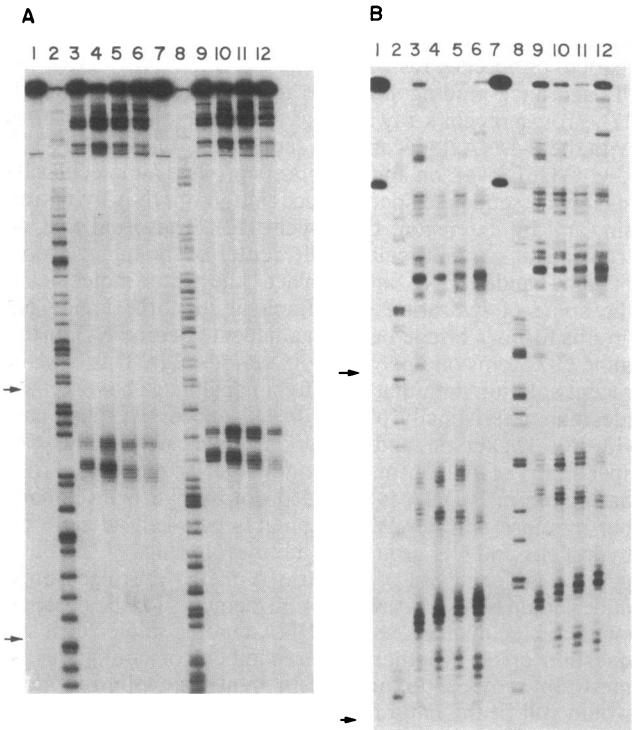

FIG. 3. Autoradiograms of high-resolution denaturing gel electrophoresis of ^{32}P-end-labeled restriction fragments (100 μM in base pairs). (*A*) The 381-base-pair (bp) *Bam*HI/*Eco*RI fragment; (*B*) the 517-bp *Eco*RI/*Rsa* I fragment. Lanes 1–6 are 3′-end-labeled DNA. Lanes 7–12 are 5′-end-labeled DNA. Lanes 1 and 7 are intact DNA control lanes. Lanes 2 and 8 contain the products of Maxam–Gilbert chemical sequencing G reactions. Lanes 3 and 9 contain products from DE·Fe(II) at 10 μM; lanes 4 and 10 contain products from P4E·Fe(II) at 5 μM; lanes 5 and 11 contain products from P5E·Fe(II); and lanes 6 and 12 contain products from P6E·Fe(II) at 3 μM. All cleavage reactions were initiated by the addition of dithiothreitol.

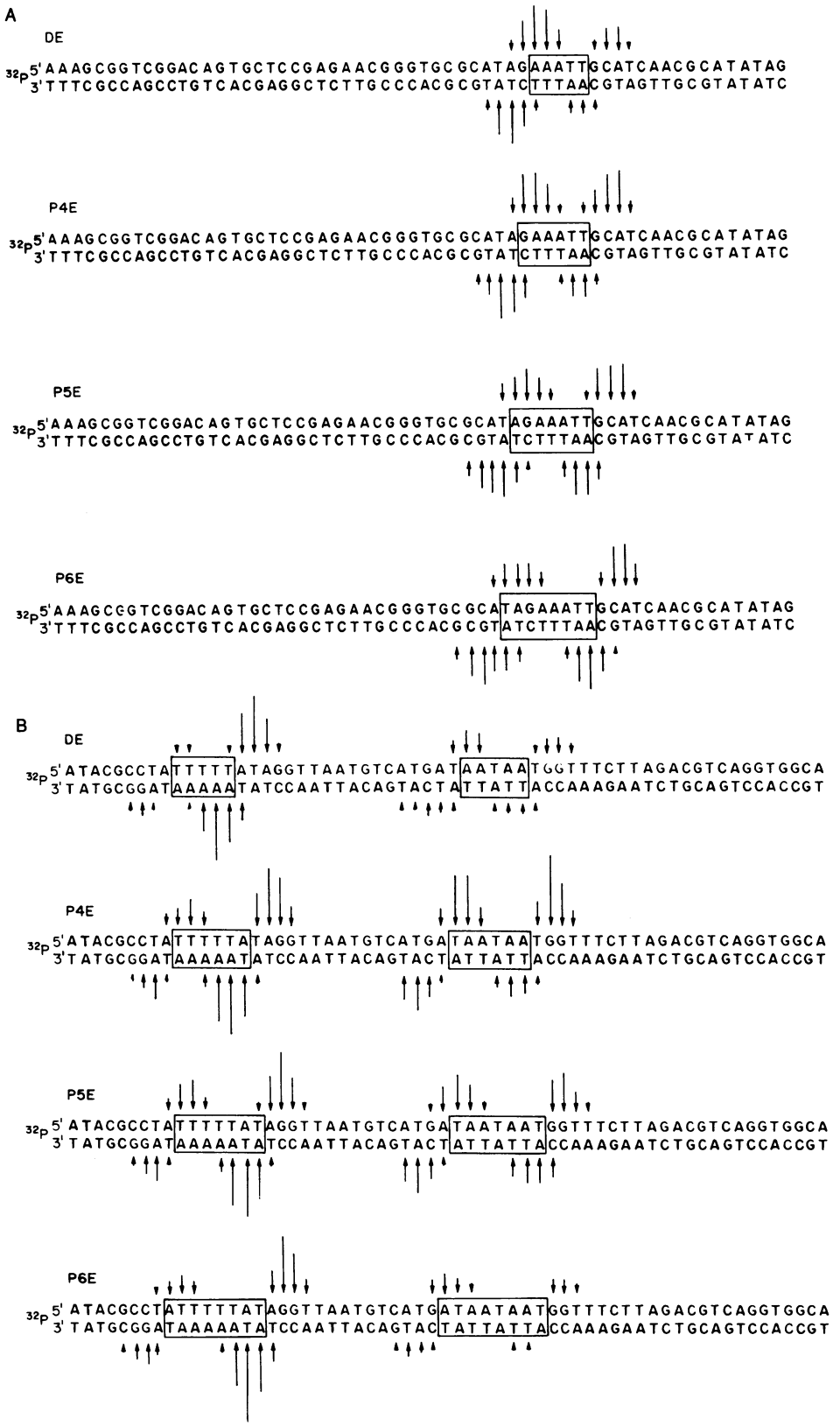

FIG. 4. Histograms of the DNA cleavage patterns on the 381-bp (pairs 229–302) (*A*) and 517-bp (pairs 4338–4278) (*B*) restriction fragments from plasmid pBR322. Boxes define binding site location and size based on Fig. 5.

terns flanking some of the A·T-rich sites are of unequal intensity.

Our interpretation of the asymmetric multiple cleavage pattern is based on a model in which the EDTA·Fe(II) moiety in a fixed or average position in the minor groove of right-handed DNA generates a diffusible short-lived reactive species, such as hydroxyl radical (15) (Fig. 5). Analysis of the DNA cleavage products shows the cleavage reaction to be oxidative degradation of the deoxyribose in the DNA backbone (20). From the known structure of right-handed B-form DNA, the deoxyriboses proximal to a fixed position of EDTA in the minor groove are on adjacent base pairs on opposite strands. From the approximate twofold rotational symmetry of the asymmetric cleavage pattern, we assign the

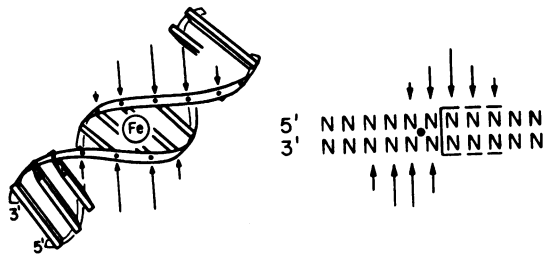

FIG. 5. Model for the asymmetric DNA cleavage pattern generated by a diffusible oxidizing species in the minor groove of right-handed DNA.

average or fixed position of the EDTA·Fe(II) moiety (Fig. 5). Because the length of the tether is known, the first base pair boundary to that oligopeptide binding site can then be identified (Fig. 5). Cleavage occurs on both sides of the A·T-rich sites, which we interpret as two roughly equi-energetic orientations of the oligopeptide-binding unit on each site. The two orientations affording two cleavage patterns flanking each site permit assignment of binding site boundaries, which are shown as boxes in Fig. 4. An assumption in this assignment is that the two orientations bind the same site rather than a fortuitous overlap of two smaller sites, each with one preferred orientation. For the tripeptides with EDTA at the amino and carboxyl termini, a *common* five-base-pair site for *both* orientations is observed (15).

381-bp Restriction Fragment. One common binding location is seen for each of the four homologous oligopeptide-EDTA reagents on the 381-bp restriction fragment (base pairs 252–259 on pBR322). There is an increase in binding site size of one base pair with each additional *N*-methylpyrrolecarboxamide unit. The orientational preference of the oligopeptide-EDTA changes with increasing chain length. The DE·Fe(II) pattern shows a bias for the tripeptide amino terminus to orient towards the 5′ end of the binding site, 5′-A-A-A-T-T-3′. The tetra- and penta-(*N*-methylpyrrolecarboxamide)s, P4E and P5E, show equi-energetic orientational preference. With P6E·Fe(II), the amino terminus of the oligopeptide orients towards the 3′ end of the sequence, 5′-T-A-G-A-A-A-T-T-3′.

517-bp Restriction Fragment. Two common binding locations are observed for the four oligopeptide-EDTA·Fe(II) molecules. There is an increase in binding site size of one base pair for each additional *N*-methylpyrrolecarboxamide unit. The orientational preference for all four oligopeptides has the amino terminus at the 3′ side of one binding site (pairs 4324–4331). There is little orientational preference at the base pair 4301–4308 site. In addition, the relative affinities of the oligopeptide change as a function of the number of oligopeptides. For DE and P6E the base pair 4331–4324 location appears to be a stronger binding site.

DISCUSSION

Binding Site Size. Four homologous oligopeptide-EDTA molecules, DE, P4E, P5E, and P6E, in the presence of Fe(II), O_2, and dithiothreitol, cleave ^{32}P-end-labeled restriction fragments at common locations rich in A·T base pairs that differ in size of the binding site. From analysis of the cleavage patterns visualized by high-resolution denaturing gel electrophoresis, the oligopeptides with three, four, five, and six *N*-methylpyrrolecarboxamide units, containing four, five, six, and seven amide NHs, bind sites of A·T-rich DNA consisting of five, six, seven, and eight contiguous base pairs, respectively (Table 1). The general rule of *n* amides affording binding site sizes of *n* + 1 base pairs is consistent with the oligopeptides binding in the minor groove of right-handed DNA with the amide NH groups forming bridges between adjacent N-3 and O-2 atoms of adenine or thymine, respectively, on opposite strands of the DNA helix.

Table 1. DNA binding sequences for the oligopeptides

DNA fragment	Oligo-peptide	*n*	Site (5′-3′)	Site size, bp
381-bp	DE	4	t-a-g-A-A-A-T-T	5
	P4E	5	t-a-G-A-A-A-T-T	6
	P5E	6	t-A-G-A-A-A-T-T	7
	P6E	7	T-A-G-A-A-A-T-T	8
517-bp	DE	4	a-t-A-A-T-A-A-t	5
	P4E	5	a-T-A-A-T-A-A-t	6
	P5E	6	a-T-A-A-T-A-A-T	7
	P6E	7	A-T-A-A-T-A-A-T	8
517-bp	DE	4	a-T-T-T-T-T-a-t	5
	P4E	5	a-T-T-T-T-T-A-t	6
	P5E	6	a-T-T-T-T-T-A-T	7
	P6E	7	A-T-T-T-T-T-A-T	8

n, Number of amide units. In each sequence, bases in the site are symbolized with capital letters and the lowercase letters show the neighboring nucleotides.

Orientation. From the relative intensities of the cleavage patterns flanking the binding site, one can estimate the relative orientational preference of the oligopeptide at each binding site 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 0.5–2.0 Å and bends back the helix axis by 8° across the region of attachment (12). One explanation for nonequivalent binding orientation on an A·T-rich binding site that lacks twofold symmetry is that the narrowness of the minor groove in B-DNA differs with local DNA sequence. The influence of flanking sequences on local structure could be important. On the 381-bp fragment, the sequence 5′-a-g-A-A-A-T-T-g-c-3′ has an orientational preference for the amino end of the tripeptide to the 5′ side. We have previously shown on a 167-bp fragment (*Eco*RI/*Rsa* I) that the same tripeptide has *no* orientational preference for the sequence 5′-t-t-A-A-A-T-T-g-c-3′ (17). Since the binding sites are identical but occur in different locations on DNA, we conclude that different local structure is conferred by flanking sequences. The oligopeptide-EDTA molecules may be useful as sensitive probes of variations in B-form DNA structure and the influence of flanking sequences.

Sequence Specificity. The literature characterizes netropsin and distamycin as molecules that bind preferably to A·T-rich regions of DNA (1–11). It is known from equilibrium binding studies of distamycin analogues that homopolymer dA·dT sequences are preferred over alternating d(A-T)·d(A-T) copolymer sequences (6, 7). The question arises whether G·C base pairs are permissible in the preferred binding sites of netropsin, distamycin, or higher oligo(*N*-methylpyrrolecarboxamide) homologs on DNA (6, 7, 23). P4E binds (5′-3′) G-A-A-A-T-T, P5E binds A-G-A-A-A-T-T, and P6E binds T-A-G-A-A-A-T-T on the 381-bp fragment (Fig. 4). Why a G·C base pair is permissible must await direct methods such as x-ray crystal structure or NMR analyses of the appropriate oligonucleotide·oligopeptide complexes. 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-rich sites that are not strong binding sites for the oligo(*N*-methylpyrrolecarboxamide)s (Fig. 4).

Design of Sequence-Specific DNA-Binding Molecules. The issue arises whether the *N*-methylpyrrolecarboxamide unit could be used as a recognition element that could be incorpo-

Table 2. Binding sequences converted to two-base-pair recognition elements

DNA fragment	Oligo-peptide	Bridged bases on opposite strands						
381-bp	DE	aa	gt	ac	AT	AT	TT	TA
	P4E	aa	gt	AC	AT	AT	TT	TA
	P5E	aa	GT	AC	AT	AT	TT	TA
	P6E	AA	GT	AC	AT	AT	TT	TA
517-bp	DE	tt	aa	AT	TT	AA	AT	tt
	P4E	tt	AA	AT	TT	AA	AT	tt
	P5E	tt	AA	AT	TT	AA	AT	TT
	P6E	TT	AA	AT	TT	AA	AT	TT
517-bp	DE	tt	TA	TA	TA	TA	aa	tt
	P4E	tt	TA	TA	TA	TA	AA	tt
	P5E	tt	TA	TA	TA	TA	AA	TT
	P6E	TT	TA	TA	TA	TA	AA	TT

Bases in the site are symbolized by capital letters.

rated into synthetic molecules designed to read B-DNA sequence specifically. According to the $n + 1$ rule and the bifurcated hydrogen bond model, the minimum recognition unit for one *N*-methylpyrrolecarboxamide on B-DNA is two base pairs. If the recognition elements for the NH carboxamide are on adjacent residues on opposite helix strands, there are 10 bridged base possibilities. AA, AT, AC, AG, TT, TC, TG, CC, GG, and CG. The data reveal that the preferences of bridged hydrogen bonds between adjacent bases on opposite helix strands decrease in the following order: AT >> AA > TT > AC, TC, TG >> AG, CG, CC, GG (Table 2). From the limited data presented here the preferred *two*-base-pair DNA sequences for the *N*-methylpyrrolecarboxamide DNA binding unit are, in decreasing order, (5′-3′)TT >> TA, AT > GA, GT, CT >> CA, CC, GC, CG.

We are grateful to the National Institutes of Health for grant support (GM-27681). This is contribution no. 7121 from the Division of Chemistry and Chemical Engineering, California Institute of Technology.

1. Hahn, F. E. (1975) in *Antibiotics III: Mechanism of Action of Antimicrobial and Antitumor Agents*, eds. Corcoran, J. W. & Hahn, F. E. (Springer, New York), pp. 79–100.
2. Zimmer, C. (1975) in *Progress in Nucleic Acids Research and Molecular Biology*, ed. Cohn, N. E. (Academic, New York), pp. 285–318.
3. Krey, A. K. (1980) in *Progress in Molecular and Subcellular Biology*, ed. Hahn, F. E. (Springer, New York), Vol. 7, p. 43.
4. McGhee, J. D. (1976) *Biopolymers* **15**, 1345–1375.
5. Van Dyke, M. W., Hertzberg, R. P. & Dervan, P. B. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5470–5474.
6. Gursky, G. V., Zasedatelev, A. S., Zhuze, A. L., Khorlin, A. A., Grokhovsky, S. L., Streltsov, S. A., Surovaya, A. N., Nikitin, S. M., Krylov, A. S., Retchinsky, V. O., Mikhailov, M. V., Beabealashvili, R. S. & Gottikh, B. P. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **74**, 367–378.
7. Krylov, A. S., Grokhovsky, S. L., Zasedatelev, A. S., Zhuze, A. L., Gursky, G. V. & Gottikh, B. P. (1979) *Nucleic Acids Res.* **6**, 289–304.
8. Luck, G., Zimmer, C., Reinert, K. E. & Arcamone, F. (1977) *Nucleic Acids Res.* **4**, 2655–2670.
9. Patel, D. J. & Canuel, L. L. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5207–5211.
10. Marky, L. A., Blumenfeld, K. S. & Breslauer, K. J. (1983) *Nucleic Acids Res.* **11**, 2857–2870.
11. Zakrzewska, K., Lavery, R. & Pullman, B. (1983) *Nucleic Acids Res.* **11**, 8825–8839.
12. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1376–1380.
13. Arcamone, F., Nicolella, V., Penco, S. & Redaelli, S. (1969) *Gazz. Chim. Ital.* **99**, 632–640.
14. Schultz, P. G., Taylor, J. S. & Dervan, P. B. (1982) *J. Am. Chem. Soc.* **104**, 6861–6863.
15. Taylor, J. S., Schultz, P. G. & Dervan, P. B. (1984) *Tetrahedron* **40**, 457–465.
16. Schultz, P. G. & Dervan, P. B. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6834–6837.
17. Schultz, P. G. & Dervan, P. B. (1984) *Biomolecular Structure and Dynamics* **1**, 1133–1147.
18. Schultz, P. G. & Dervan, P. B. (1983) *J. Am. Chem. Soc.* **105**, 7748–7750.
19. Hertzberg, R. P. & Dervan, P. B. (1982) *J. Am. Chem. Soc.* **104**, 313–315.
20. Hertzberg, R. P. & Dervan, P. B. (1984) *Biochemistry* **23**, 3934–3945.
21. Sutcliffe, J. G. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77–90.
22. Peden, K. (1983) *Gene* **22**, 277–280.
23. Zimmer, C., Luck, G., Birch-Hirschfeld, E., Weiss, R., Arcamone, F. & Guschlbauer, W. (1983) *Biochim. Biophys. Acta* **741**, 15–22.