Design of Sequence-Specific DNA-Binding Molecules

Peter B. Dervan

Science, New Series, Vol. 232, No. 4749. (Apr. 25, 1986), pp. 464-471.

Stable URL:
http://links.jstor.org/sici?sici=0036-8075%2819860425%293%3A232%3A4749%3C464%3ADOSDM%3E2.0.CO%3B2-7

Science is currently published by American Association for the Advancement of Science.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/aaas.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org/
Wed Mar 1 12:03:14 2006

mothers who personally take care of their children would encourage women to stay out of the labor force. Some policies would tend to lower fertility, others would raise it. Some would discourage employers from hiring women while others would not. A major challenge to social science research is to understand the economic and social consequences of alternative policies.

Although there is disagreement on theory and policy, virtually everyone agrees on one point. As long as parents are responsible for children and this responsibility is borne disproportionately by women, sex differences in the labor market are likely to persist.

REFERENCES AND NOTES

1. G. Cain, in *Handbook of Labor Economics*, O. Ashenfelter and R. Layard, Eds. (North-Holland, Amsterdam, in press).
2. V. R. Fuchs, *The Service Economy* (National Bureau of Economic Research, New York, 1968).
3. J. Mincer, in *Aspects of Labor Economics* (Princeton Univ. Press, Princeton, 1962), pp. 63–97.
4. R. Easterlin, *Birth and Fortune: The Impact of Numbers on Personal Welfare* (Basic Books, New York, 1980).
5. B. Friedan, *The Feminine Mystique* (Norton, New York, 1963).
6. U.S. Department of Commerce, Bureau of the Census, *Census of Population and Housing, Public Use Samples: One-in-One-Thousand Samples* (Washington, DC, 1960, 1970, and 1980); *Current Population Survey: Annual Demographic File* (Washington, DC, 1980 and 1984).
7. A check of this method for 1969 and 1979 (2 years when both Census and CPS data are available) revealed that the rates of change of earnings are similar in both sources.
8. Institute for Social Research, *Time Use in Economic and Social Accounts, 1975–1976* (University of Michigan, Ann Arbor, 1978).
9. J. P. Robinson, in *Time, Goods, and Well-Being*, F. T. Juster and F. P. Stafford, Eds. (Institute for Social Research, University of Michigan, Ann Arbor, in press).
10. M. A. Ferber and C. A. Green, *J. Hum. Resour.* **20**, 90 (1985).
11. E. Lazear and R. T. Michael, *Am. Econ. Rev.* **70**, 91 (1980).
12. ——, *Allocation of Income Within the Household* (Univ. of Chicago Press, Chicago, in press).
13. G. S. Becker, *A Treatise on the Family* (Harvard Univ. Press, Cambridge, MA, 1981).
14. V. R. Fuchs, "His and hers: Gender differences in work and income, 1959–1979" (NBER working paper 1501, National Bureau of Economic Research, Cambridge, MA, 1984); *J. Labor Econ.*, in press.
15. K. Mason, in *Women and the Workplace*, M. Blaxall and B. Reagan, Eds. (Univ. of Chicago Press, Chicago, 1976), p. 81.
16. G. S. Becker, *J. Labor Econ.* **3**, S33 (1985).
17. J. J. Heckman, *Econometrica* **47**, 153 (1979).
18. J. P. Smith and M. P. Ward, *Women's Wages and Work in the 20th Century* (Rand Corporation, Santa Monica, CA, 1984).
19. Calculated by the author from *Current Population Survey* tapes.

20. Financial support for this work was provided by the Alfred P. Sloan Foundation and the Rockefeller Foundation. The research assistance of Leslie Perreault, and comments from John Pencavel and Joyce Jacobsen are also gratefully acknowledged.

Design of Sequence-Specific DNA-Binding Molecules

PETER B. DERVAN

Base sequence information can be stored in the local structure of right-handed double-helical DNA (B-DNA). The question arises as to whether a set of rules for the three-dimensional readout of the B-DNA helix can be developed. This would allow the design of synthetic molecules that bind DNA of any specific sequence and site size. There are four stages of development for each new synthetic sequence-specific DNA-binding molecule: design, synthesis, testing for sequence specificity, and reevaluation of the design. This approach has produced bis(distamycin)fumaramide, a synthetic, crescent-shaped oligopeptide that binds nine contiguous adenine-thymine base pairs in the minor groove of double-helical DNA.

X-RAY ANALYSIS OF CRYSTALS OF RIGHT-HANDED DOUBLE-helical DNA (B-DNA) reveals that base sequence information can be stored in the local structure of the helix (*1, 2*). The DNA polymer consists of GC and AT base pairs like rungs on a twisted ladder. The helical twist, groove shape, base pair slide and roll, and handedness of double-helical DNA depend on base sequence (*1–7*). In this article I address the issue of whether a set of rules for the three-dimensional readout of B-DNA could be developed based on low molecular weight natural products commonly identified as antibiotic, antiviral, or anticancer DNA-binding drugs. This would allow the design of synthetic nonprotein molecules that bind B-DNA of any base sequence and site size. The tools of synthetic and mechanistic 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.

There are four bases possible for each nucleotide position on each strand of the DNA helix, and, within the constraints of the AT and GC 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 (Table 1). During the past few years, a priority has been to develop the analytical methods needed to analyze precisely the sequence specificities of either natural or synthetic DNA-binding small molecules. These methods are footprinting (*8–18*) and affinity cleaving (*19–25*). With regard to the overall experimental approach, there are four stages of development for each new synthetic DNA-binding molecule: design, synthesis, testing for sequence specificity, and reanalysis of the design.

The initial design ideas are derived from examination of natural

P. B. Dervan is professor of chemistry at California Institute of Technology, Pasadena, CA 91125.

products that bind to specific sequences in the minor groove of B-DNA (*26*). Some of these DNA-binding molecules are flat, and this shape allows them to sandwich or intercalate between the base pairs. Other natural products with less obvious structural features are believed to fit snugly in the minor or major grooves of the right-handed DNA helix by hydrogen bonds, electrostatic interactions, and van der Waals interactions. These natural products are in the molecular weight range of 500 to 2000 and are sufficiently large to cover two to five contiguous base pairs, up to one half-turn of the DNA helix. Binding-site sizes of two to five base pairs have 10 to 512 unique combinations of base pairs or specific binding sites on double-helical DNA, respectively (Table 1). What is the combination of nonbonded stabilizing and destabilizing interactions that allows structurally diverse natural products such as netropsin and distamycin to bind AT-rich sequences of DNA and echinomycin, triostin A, bleomycin, actinomycin D, and chromomycin to bind GC-rich sequences of DNA? X-ray structures are the most reliable sources for model building. A small but significant number of crystal structures of small molecule-oligonucleotide complexes such as actinomycin D (*27, 28*), daunomycin (*29*), triostin A (*30*), echinomycin (*31*), and netropsin structures (*32*) 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 (*33*).

Table 1. Relation between binding-site size and number of distinguishable sequences.

n*	Unique sites (number)
2	10
3	32
4	136
5	512
6	2,080
7	8,192
8	32,896
9	131,072
10	524,800

*Site size (in base pairs).

This design-synthesis-footprinting (or affinity cleaving) exercise is relevant to the general problem of refining our understanding of molecular recognition. Within the area of nucleic acids, the development of synthetic molecules that can read large sequences of double-helical DNA may lead to the development of new research tools for use in molecular biology, diagnosis of disease states at the level of DNA, and novel chemotherapeutic strategies.

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 (*8*) or the synthetic reagent methidiumpropyl-EDTA (MPE) (*9, 34, 35*), 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 base pair 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. This footprinting technique was first described by Galas and Schmitz, who used DNase I to determine the sequence preferences of DNA-binding proteins that cover about 20 base pairs, or approximately two turns of the DNA helix (*8*).

For smaller molecules (500 to 2000 molecular weight) such as the antibiotic, antiviral, and antitumor drugs, footprinting with DNase I reveals binding-site sizes that are several base pairs larger than would be expected from model building of these molecules complexed to DNA (*11, 12*). Because binding-site size is a useful parameter for defining the sequence specificities of small molecules on DNA, a synthetic footprinting tool, called MPE · Fe(II), is often employed because it mimics DNase I as a DNA-cleaving reagent and affords accurate resolution of the binding-site sizes for DNA binding small molecules in footprinting experiments.

MPE contains the DNA intercalator methidium covalently bound by a short hydrocarbon tether to the metal chelator EDTA (*34, 35*) (Fig. 1). In the presence of ferrous ion, reducing agents such as dithiothreitol and dioxygen, MPE at micromolar concentrations produces single-strand breaks at 25°C (*p*H 7.0) in double-helical DNA (*34, 35*). The synthetic MPE · Fe(II) cleaves DNA with lower sequence specificity than DNase I (*14, 15*). By means of MPE · Fe(II) footprinting, the binding locations and site sizes of several natural products such as actinomycin D, distamycin, netropsin, chromomycin, mithramycin, olivomycin, and echinomycin on DNA restriction fragments have been determined (*9, 10, 13–15, 18*) (Fig. 2).

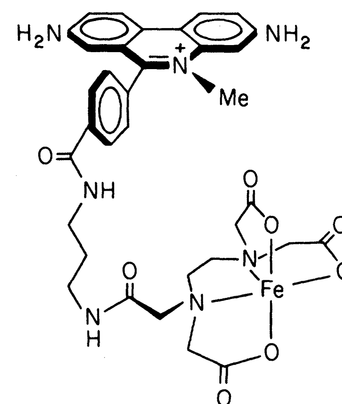

Fig. 1. Methidiumpropyl-EDTA · Fe(II) [MPE · Fe(II)]. Me, methyl group.

X-ray crystal structures reveal that DNA accommodates some of these small molecules by changing its shape and conformation (*29–32*). For example, the crystal structures of echinomycin and triostin A complexed to the DNA fragment with sequence (5′-3′)-CGTACG reveal that the bisintercalator reorganizes the nucleic acid helix to a novel structure (*30, 31*). The bisintercalators create a Hoogsteen base pair at an A · T base pair flanking the quinoxaline-intercalating moiety with A in the *syn* conformation. In addition,

base pairs flanking the intercalation site are unwound. How far from the intercalation 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 of investigators have observed enhanced rates of cleavage flanking the binding sites of the antibiotics distamycin, actinomycin, and echinomycin (*11, 12, 14–17*). This enhanced susceptibility to cleavage by DNase I has been interpreted by Waring and co-workers as the antibiotic altering the width of the minor groove at nearby sequences (*17*). 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 at 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 are distributed over neighboring regions of DNA—will influence our choice of coupling strategies for synthetic hybrids of different DNA-binding natural products.

Affinity Cleaving

With the finding that attachment of EDTA · Fe(II) to a DNA-binding molecule such as methidium creates an efficient DNA-cleaving molecule (at 25°C and *p*H 7.0), attachment of EDTA to sequence-specific DNA-binding molecules was examined (*19, 20*). The EDTA attachment converts a sequence-specific DNA-binding molecule to a sequence-specific DNA-cleaving molecule (*19, 20*). Analysis of the cleavage products from a ^{32}P–end-labeled restriction fragment on the autoradiogram of a high-resolution denaturing polyacrylamide gel allows the binding locations, site size, and orientation of synthetic molecules on double-helical DNA to be visualized (*19, 20*). The resulting cleavage patterns are the positive image visualized on an autoradiogram with respect to the negative image produced by footprinting (Fig. 3).

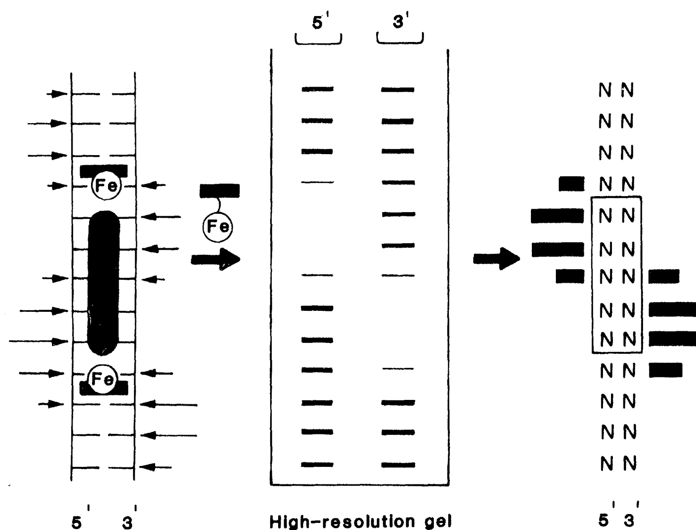

Fig. 2. Scheme for the MPE · Fe(II) footprinting method. (Left) Sequence-specific DNA-binding molecule protects a region of a restriction fragment labeled with ^{32}P at the 5′ (and 3′) end from oxidative cleavage of the DNA backbone generated by nearby bound MPE · Fe(II) molecules. The arrows indicate cleavage intensities on opposite DNA strands resulting from nearest bound MPE · Fe(II). Asymmetric cleavage protection to the 3′ side of each strand is due to a diffusible reactive species (most likely hydroxyl radical) capable of cleaving DNA generated in the minor groove of B-DNA (*20, 35*). (Center) Autoradiogram, which allows visualization of cleavage patterns resulting from DNA cleavage products on a high-resolution denaturing gel. (Right) Histograms represent missing DNA cleavage products resulting from protection from MPE · Fe(II) cleavage. Box represents an assignment of binding locations and site size.

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 AT-rich sequences (*36–44*). 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 (*20*) (Fig. 4). DE · Fe(II) and ED · Fe(II) at micromolar concentrations cleave pBR322 plasmid DNA at discrete locations 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 AT-rich DNA. The multiple contiguous cleavages at each site are taken as evidence for a diffusible oxidizing species, most likely hydroxyl radical (*20, 35*). 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. A comparison of ED · Fe(II) with DE · Fe(II) shows DNA-cleavage patterns flanking the same site (for example, 5′–AAATT–3′) but with intensities of opposite polarities (*20*). The cleavage patterns on opposite DNA strands are asymmetric, that is, shifted to the 3′ side, which can be understood by examination of a model of B-DNA (Fig. 5). 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. A model for the relation between the cleavage pattern and the binding site is given in Fig. 5. If the major and minor orientations of DE · Fe(II) bind the same site, it appears that the tripeptide unit of distamycin covers 5 bp. Analysis of the ED · Fe(II) cleavage pattern confirms this conclusion. The solution data support a model where each of the four pyrrolecarboxyamide NH's can hydrogen-bond to two bases on the floor of the minor groove on adjacent base pairs on opposite strands of the helix (*21, 32*) (Fig. 6).

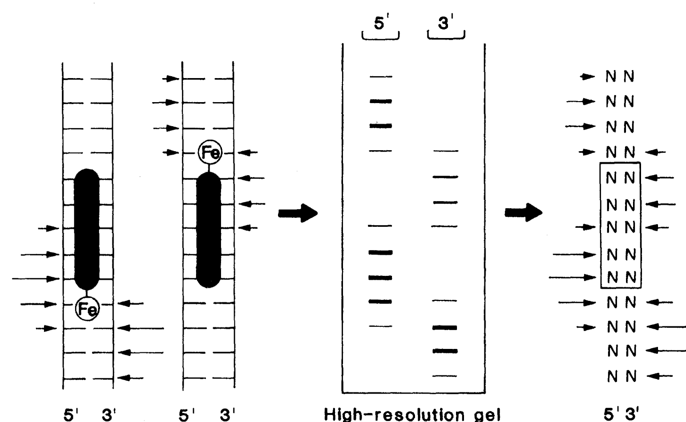

Fig. 3. Scheme for the affinity-cleaving method. (Left) Sequence-specific DNA-binding molecule equipped with EDTA · Fe(II) can assume two orientations, affording asymmetric cleavage patterns on opposite strands of a restriction fragment labeled with ^{32}P at the 5′ (and 3′) end. Lengths of arrows indicate frequency of cleavage. (Center) Autoradiogram of cleavage patterns of a high-resolution denaturing gel. (Right) Assignment of binding-site boundaries based on the model in Fig. 4.

Recognition of Large Contiguous Sequences of AT-Rich DNA

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 (*32*). 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 adjacent 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 AT-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 (*45, 46*). Tetra-, penta-, and hexa(*N*-methylpyrrolecarboxamide)s equipped with EDTA (P4E, P5E, and P6E) were synthesized and compared on sequencing gels by means of affinity cleaving (*23*) (Fig. 7).

The homologous oligopeptide-EDTA · Fe(II) molecules cleave restriction fragments at common locations rich in AT that differ incrementally in the size of the binding site (*23*). From analysis of the cleavage patterns visualized by high-resolution denaturing gel electrophoresis, the oligopeptides with four, five, and six *N*-methylpyrrolecarboxamide units and containing five, six, and seven amide groups bind sites of AT-rich DNA consisting of six, seven, and eight contiguous base pairs, respectively (*23*) (Table 2). 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 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 (*32*). 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′-*AG*AAATT*GC*-3′. DE has no orientation preference for the sequence 5′-*TT*AAATT*GC*-3′. Since the binding sites are the same but occur in different locations on DNA, we conclude that different local structures can be conferred on identical binding sites by flanking sequences.

Table 2. Comparison of DNA-binding sequences for synthetic oligopeptides. In each sequence bases in the binding site are symbolized by capital letters, and the italic letters show the neighboring nucleotides. Data taken from one major binding location on a 517-bp restriction fragment from pBR322 DNA (pairs 4338–4278) (*23, 25*).

Oligo-peptide	n*	Site (5′-3′)	Site size (base pairs)
DE	4	*A*TTTTT*ATA*	5
P4E	5	*A*TTTTTA*TA*	6
P5E	6	*A*TTTTTAT*A*	7
P6E	7	ATTTTTAT*A*	8
BEDF	8	ATTTTTATA	9

*Number of amide units.

Netropsin and distamycin have been characterized as molecules that bind preferably to AT-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 GC base pairs are permissible in the preferred binding sites of netropsin, distamycin, and higher oligo(*N*-methylpyrrolecarboxamide) homologs on DNA (*18, 23*). For example, P4E binds (5′-3′)GAAATT, P5E binds AGAAATT, and P6E binds TAGAAATT (*23*). 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,

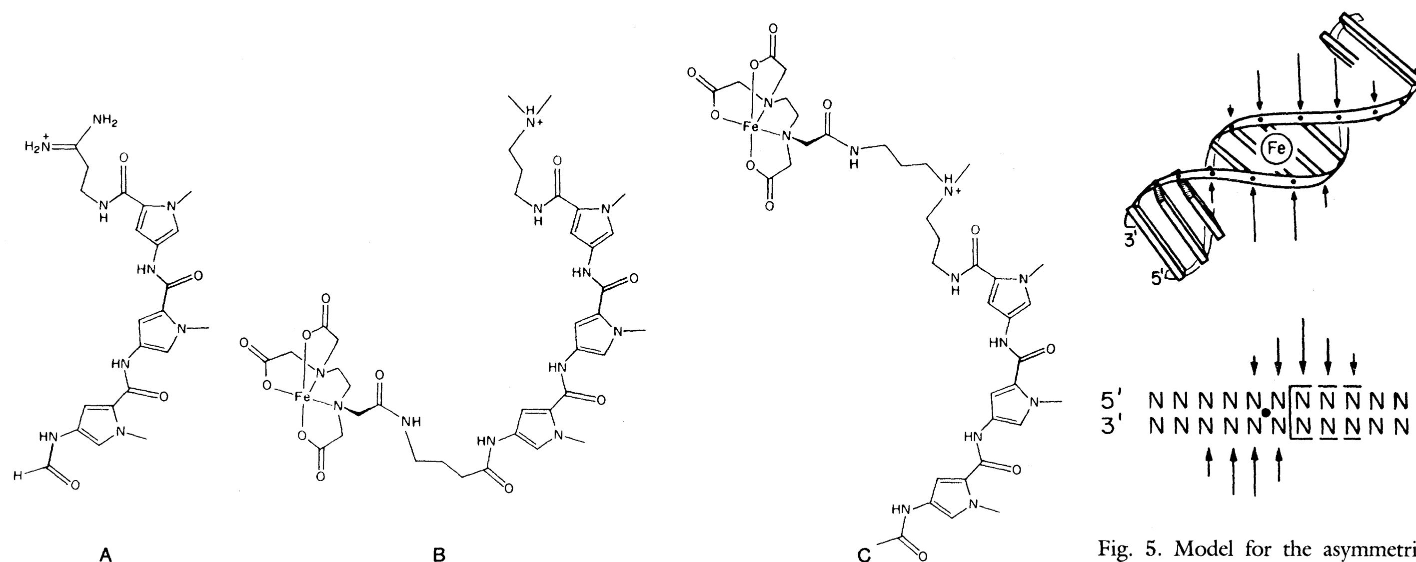

Fig. 4. (A) Distamycin, (B) distamycin-EDTA · Fe(II) [DE · Fe (II)], and (C) EDTA-distamycin · Fe(II) [ED · Fe(II)].

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

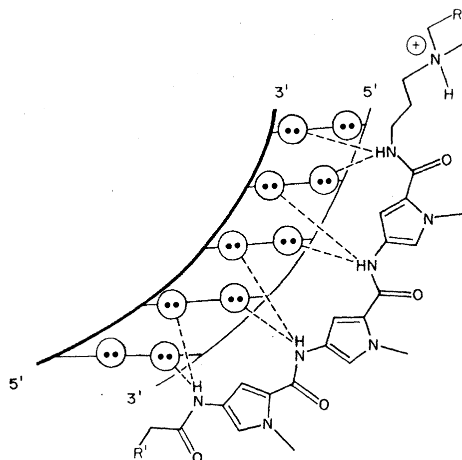

Fig. 6. Model for *tris*-*N*-methylpyrrolecarboxamide binding in the minor groove of DNA at AT-rich sequences 5 bp in size, based on affinity cleaving data in solution (*21*) and x-ray crystal structure of netropsin:DNA in the solid state (*32*). Circles with two dots represent lone pairs of electrons on N-3 of adenine and O-2 of thymine at the edge of the base pairs on the floor of the minor groove of B-DNA. Dotted lines are bridged hydrogen bonds to the amide NH's. For DE, R = $(CH_2)_2NHCOCH_2N(CH_2CO_2H)$ $CH_2CH_2N(CH_2CO_2H)_2$ and R' = H. For ED, R = H and R' = $(CH_2)_2NHCOCH_2N(CH_2CO_2H)$ $CH_2CH_2N(CH_2CO_2H)_2$.

there are several pure AT 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. The preferred 2-bp 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 (*23*).

Coupled DNA Binding Units of Similar and Mixed Specificity

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 very 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 (*24, 25, 47*). For the success of this coupled DNA-binding unit strategy, the base-specific recognition elements of each subunit and the linkers connecting them must be compatible with the same groove (major or minor in B-DNA) and conformational state of the DNA.

The initial design attempts for large sequences of AT-rich DNA involved the construction of dimers of di- and tripeptides connected by flexible hydrocarbon tethers (*24, 47*). Khorlin and co-workers constructed two netropsin-like molecules linked by C_4, C_6, C_8, and C_{12} hydrocarbon biscarboxamide linkers (*47*). From saturation levels of binding, they concluded that the dimer of netropsin with a C_8 spacer afforded simultaneous binding of both dipeptides (*47*). From affinity cleaving experiments, we found that a tripeptide dimer with the C_7 linker bis(EDTA-distamycin)Fe(II) [BED · Fe(II)] binds a 9-bp AT-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 (*24*).

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 (Fig. 8). A second molecule, 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 (Fig. 9), reveals major cleavage sites flanking two AT-rich sequences, 5'-ATTTTTATA-3' and 5'-ATAATAAT-3' (*25*). The observation of exclusive 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.

Several natural products that sequence-specifically 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 (*27, 28*). Actinomycin is an intercalator that unwinds DNA by 26° (*48*) and binds 4 bp with a preference for 5'-NGCN-3' sequences (*9–12*). The cyclic pentapeptides fit snugly above and below the intercalating ring in the minor groove of DNA.

X-ray crystal structures for 1:2 complexes of actinomycin with d(GpC) and dG show an intercalative structure and specific hydrogen bonds from the guanine 2-amino group to the carbonyl oxygen of the L-threonine residue as well as two hydrogen bonds between the D-valine residues of the peptide rings (*27, 28*). As determined from footprinting data, the binding-site locations of both the AT-binding distamycin and the GC-binding actinomycin are known on several DNA restriction fragments (*9–18*). From DNase I footprinting experiments, enhancement of DNA cleavage contiguous to sites

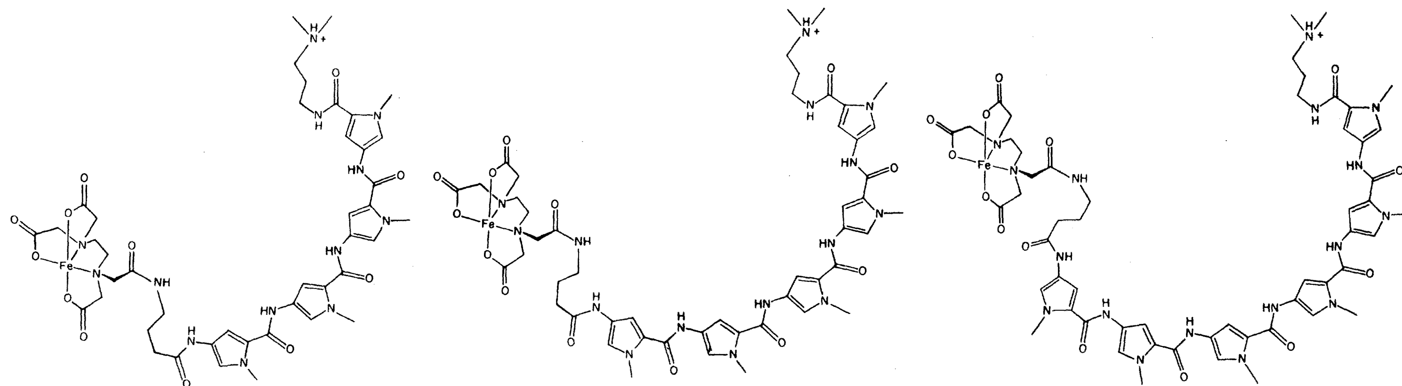

Fig. 7. Tetra-, penta-, and hexa(*N*-methylpyrrolecarboxamide)-EDTA · Fe(II) [P4E · Fe(II), P5E · Fe(II), P6E · Fe(II)].

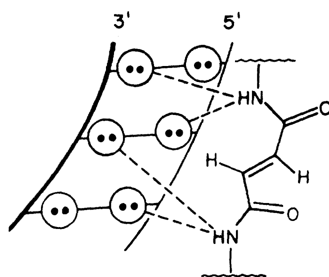

Fig. 8. Model of fumaramide binding in the minor groove of AT-rich DNA.

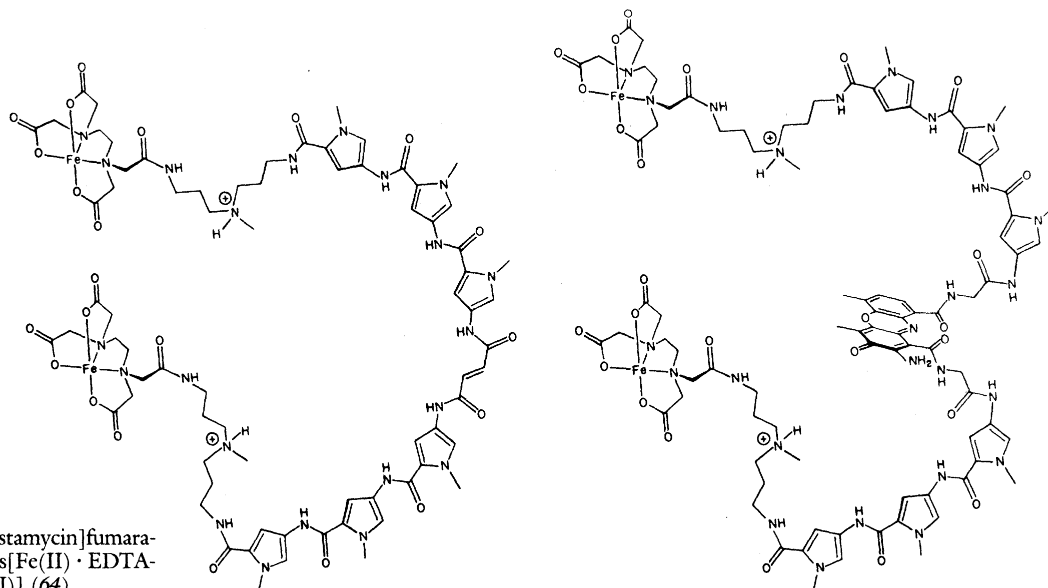

Fig. 9. (Left) Bis[Fe(II) · EDTA-distamycin]fumaramide [BEDF · Fe(II)] (*25*). (Right) Bis[Fe(II) · EDTA-distamycin]phenoxazone [BEDP · Fe(II)] (*64*).

bound by actinomycin afford some idea of the extent of local helix distortion caused by binding of actinomycin to double-helical DNA. Waring found that this ligand-induced effect can be propagated in either direction from the binding site over more than one turn of the helix (*17*). Not every bound molecule produces an enhancement of cleavage at flanking sequences.

Given the different binding modes of actinomycin and distamycin (intercalation compared to minor groove binding), the question arises as to whether the DNA helix can accommodate both these ligands within close proximity, despite their unique distortions, to the DNA structure. There is evidence indicating that proximal binding of the two ligands is possible. Spectral titration, circular dichroism, and methylation studies of the effect of actinomycin binding to distamycin-DNA complexes indicate that binding of actinomycin does not give rise to appreciable dissociation of distamycin at normal binding densities (*27*). Spectral titration data for the simultaneous binding of actinomycin and netropsin indicate that adjacent binding is possible (*49*). A nuclear magnetic resonance study of simultaneous netropsin and actinomycin binding to a synthetic dodecamer (5′-CGCGAATTCGCG-3′) provides further evidence for adjacent binding (*50*).

Because simultaneous intercalation of actinomycin and minor groove binding of distamycin in close proximity are possible, the question arises as to whether an intercalator and a minor groove binder can be covalently linked in such a way that the sequence specificity and the ability to bind both moieties simultaneously is retained. Krivtsova and co-workers have reported the synthesis of a series of hybrid molecules, called distactins, that have one, two, and three *N*-methylpyrrolecarboxamides directly attached to the 1,9 positions of the phenoxazone chromophore (*51*). From spectrophotometry, viscometry, and flow birefringence studies, they concluded that the distactin containing three *N*-methylpyrrole rings is bound to DNA only on the outside of the double helix through the peptide fragments and that intercalation of the phenoxazone does not take place (*51*). Distactins with one or two *N*-methylpyrrole rings appear to have both intercalation and groove binding modes (*51*).

In our laboratories, the two cyclic pentapeptides of actinomycin were replaced with the tripeptide from distamycin and, in addition, the postulated key carbonyl amide guanine NH recognition element of actinomycin was retained (Fig. 9). As judged from distance and steric considerations from model building studies, the distamycin tripeptide was connected to the phenoxazone by a glycine tether (*52*) (Fig. 9). The resulting bis(distamycin)phenoxazone is a possible groove binder–intercalator–groove binder that may bind 10 bp of DNA having the sequence $(AT)_4GC_2(AT)_4$ (Fig. 10). The sequence specificity of bis(distamycin)phenoxazone on DNA restriction fragments was tested by the affinity cleaving method, which required attachment of the DNA-cleaving function EDTA-Fe(II) (*52*).

The DNA cleavage patterns reveal a major cleavage site flanking the 10-bp sequence, 5′-TATAGGTTAA-3′ (*52*). One interpretation of the data is that the tripeptides are binding simultaneously at AT-rich sequences 4 bp in size flanking the central 5′-GC-3′ phenoxazone binding site. Although not a proof, this is 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. Perhaps at these sites there is intercalation of the phenoxazone with distortions on both sides of the intercalation site but with the local distortion being sequence dependent, which would make one site incompatible with the tripeptide groove binder. Because there are 524,800 unique sequences of double-helical DNA that are 10 bp

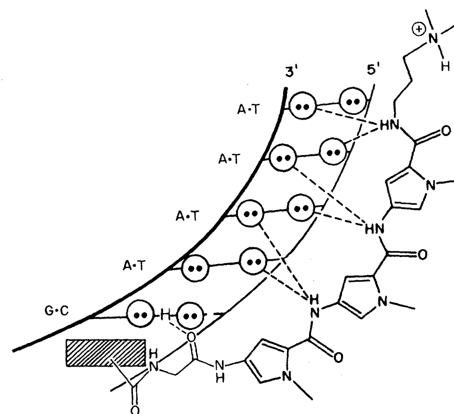

Fig. 10. Model for a hybrid intercalator–groove binder, distamycin-phenoxazone binding to 5 bp $(GC)(AT)_4$ of DNA. Circles with two dots represent lone pairs of electrons on N-3 of adenine (A) and O-2 of thymine (T). Circles with two dots and H interposed represent lone pair of electrons on N-3 of guanine (G) and O-2 of cytosine (C) flanking the -NH_2 of guanine (G) at the edges of the base pairs on the floor of the minor groove of right-handed DNA. Dotted lines are hydrogen bonds.

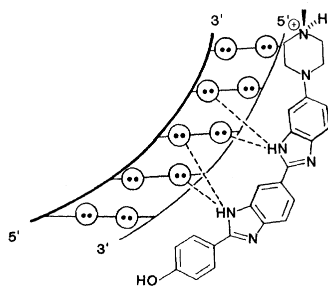

Fig. 11. Model for Hoechst 33258 in the minor groove of B-DNA. This is a refinement of the Mikhailov model (*53*) and is based on the x-ray structure of netropsin:DNA (*32*). Circles with two dots represent lone pairs of electrons on N-3 of adenine and O-2 of thymine at the edges of the base pairs on the floor of the minor groove of the DNA helix. Dotted lines are bridged hydrogen bonds to the bisbenzimidazole NH.

in size; it is likely that we have not yet identified the optimal 10-bp recognition site for bis(distamycin)phenoxazone.

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 AT words can be coupled with GC words. From a comparison with the distactins, the choice of a linker connecting the recognition elements may be a critical design feature with regard to simultaneous binding of all moieties.

General Features of AT Recognition in the Minor Groove

Hoechst dye 33258 is a bisbenzimidazole that binds double-helical DNA rich in A and T (*53, 54*). Martin and Holmes have shown that the binding of Hoechst 33258 on DNA requires a minimum of four consecutive AT base pairs (*54*). Hoechst is similar in size and shape to netropsin and distamycin. The inside edge of the crescent framework of Hoechst has potential NH recognition elements similar distance in space to the carboxamide NH's of netropsin and distamycin. Mikhailov and co-workers have suggested that the bisbenzimidazole framework of Hoechst binds to B-DNA, similar to netropsin and distamycin (*53*). Footprinting methods were used to determine whether Hoechst, netropsin, and distamycin, molecules similar in shape but different in chemical structure, share common binding sites on B-DNA (*18*).

Indeed, Hoechst, netropsin, and distamycin share common binding sites that are 5 ± 1 bp in size and rich in AT base pairs (*18*). The 5-bp protection patterns for Hoechst 33258 may result from a central 3-bp recognition site bound by two bisbenzimidazole NH's forming a bridge on the floor of the minor groove between adjacent adenine N-3 and thymine O-2 atoms on opposite helix strands. Hydrophobic interaction of the flanking phenol and *N*-methylpiperazine rings with the walls of the minor groove would afford a steric blockade of one additional base pair on each side (Fig. 11). The correspondence of similar binding locations of Hoechst 33258, netropsin, and distamycin structures suggests that flat aromatic rings twisted in a screw sense to match the walls of the DNA helix and oriented on the floor of the helix by one or more bridged hydrogen bonds may be a general feature of recognition at AT-rich sequences of double-helical DNA (*18*).

Alteration of the Sequence Specificities of Natural Products

So far I have discussed coupling sequence-specific DNA-binding subunits derived from low molecular weight natural products, an effort directed toward recognition of larger binding site sizes and hence higher sequence specificity on DNA. A strategy of altering natural products to change their sequence specificities is another possibility. There is one example of this from Olsen's laboratories, in which a natural product bisintercalator that binds 5′-NCGN-3′ was converted to a modified synthetic version that binds 4 bp of AT-rich DNA (*55–57*).

The quinoxaline antibiotics echinomycin and triostin A are natural products that consist of a cross-bridged cyclic octadipeptide bearing two quinoxaline chromophores (*56*). The two planar rings of echinomycin and triostin A bisintercalate in the minor groove of the DNA double helix (*58*). As determined from footprinting studies, the structurally similar echinomycin and triostin A have a binding-site size of 4 bp that contains the central 2-bp sequence 5′-CG-3′ (*15, 16*). The three-dimensional structure of a complex between triostin A and a DNA duplex, 5′-CGTACG-3′, has been solved by Rich and co-workers (*30*). They find direct evidence that triostin A is a bisintercalator that brackets the 2-bp sandwich 5′-CG-3′, forming three hydrogen bonds between the L-alanine of the octapeptide backbone of triostin A and the guanine in the minor groove of DNA. The NH groups of both alanine residues on triostin A form hydrogen bonds to the guanine on opposite and adjacent base-pair residues of the DNA duplex. However, only one carbonyl of the two alanines forms a hydrogen bond to guanine, suggesting that recognition of triostin A on opposite strands of DNA is unequal. In addition, as mentioned in the introduction, these investigators made the remarkable observation that the AT base pairs on the outside of the intercalation site are rearranged from Watson-Crick to Hoogsteen pairing (*30*).

Olsen has synthesized a triostin A analog in which all four *N*-methyl groups in the natural product are absent (*55*). This des-*N*-tetramethyl triostin A, called TANDEM, binds preferentially to alternating AT sequences, a change in specificity from the natural product triostin A (*57*). Perhaps the absence of the *N*-methyl groups leads to intramolecular hydrogen bonding between the NH amides of TANDEM and the lone pairs of electrons on N-3 of adenine and O-2 of thymine on the floor of the minor groove of AT-rich DNA.

Whether this example of altering natural product sequence specificities can be repeated in other systems remains to be seen. In the netropsin-distamycin series, Dickerson and Lown have made the suggestion that replacement of pyrrole in the natural product by imidazole may generate a netropsin analog capable of accepting GC sequences in preference to AT base pairs (*32*).

Reflections and Conclusion

It is fair to say that the area of molecular recognition of DNA and the design of sequence-specific DNA-binding molecules is in its infancy. With the availability of analytical techniques such as footprinting and affinity cleaving, which allow rapid and precise analysis of hundreds of potential binding sites on large DNA, one may be optimistic for further progress in this area. Work to date has focused on recognition in the minor groove of B-DNA, which is a consequence of the fact that the modeled natural products bind in the minor groove of right-handed DNA. This is in contrast to sequence-specific recognition of DNA by proteins that appear to bind largely in the major groove of DNA (*59–64*). There is reason to expect that some of the chemical principles that are uncovered in the minor groove approach, such as the role of complementary hydrogen bonding between the edges of the base pairs on the floors of the DNA helix, will be applicable to the major groove recognition problem. In fact, the major groove is much less symmetric with regard to donor-acceptor relations on the floor of the helix groove and should be the preferred target for DNA-binding molecules of very high or absolute sequence specificity (*65*).

Although there has been some encouraging success with regard to building synthetic molecules that bind large sequences of AT-rich double-helical DNA, there has not been corresponding success in the development of well-understood GC recognition elements. Progress in this area is an important component in an overall strategy of coupling GC and AT words into sentences that uniquely recognize long sequences of right-handed DNA. Moreover, from x-ray crystal structural work, we are now gaining an appreciation for the plasticity of double-helical DNA. Until we understand how the various helix shapes, conformations, local sugar pucker, and backbone torsion are altered to accommodate such a diverse structural group of DNA-binding molecules, we will be limited to model building and designing synthetic targets that are complementary to the classic B-DNA structure (*1*).

Finally, there seems to be lack of new low molecular weight natural products that are sequence-specific DNA-binding molecules. The natural products netropsin, distamycin, bleomycin, actinomycin, echinomycin, and chromomycin have been known for decades or more. Perhaps the next generation of sequence-specific DNA-binding molecules that will come under the scrutiny of biophysical techniques, x-ray crystal structural analyses, calculational methods, and medicinal interest will be synthetic. Therefore, the organic chemist is bound to become a key partner with the crystallographer, the biophysical chemist, and the theoretician in solving this challenging chemical problem.

REFERENCES AND NOTES

1. R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini, M. L. Kopka, *Science* **216**, 475 (1982).
2. R. E. Dickerson, *J. Mol. Biol.* **166**, 419 (1983).
3. O. Kennard, *Pure Appl. Chem.* **56**, 989 (1984).
4. A. H-J. Wang *et al.*, *Nature (London)* **282**, 680 (1979).
5. A. Rich, A. Nordheim, A. H-J. Wang, *Annu. Rev. Biochem.* **53**, 791 (1984).
6. C. R. Calladine, *J. Mol. Biol.* **161**, 343 (1982).
7. ——— and H. R. Drew, *ibid.* **178**, 773 (1984).
8. D. J. Galas and A. Schmitz, *Nucleic Acids Res.* **5**, 3157 (1978).
9. M. W. Van Dyke, R. P. Hertzberg, P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5470 (1982).
10. M. W. Van Dyke and P. B. Dervan, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 347 (1982).
11. A. V. Scamrov and R. Sh. Beabealashvilli, *FEBS Lett.* **164**, 97 (1983).
12. M. Lane, J. C. Dabrowiak, J. N. Vournakis, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3260 (1983).
13. M. M. Van Dyke and P. B. Dervan, *Biochemistry* **22**, 2373 (1983).
14. ———, *Nucleic Acids Res.* **11**, 5555 (1983).
15. ———, *Science* **225**, 1122 (1984).
16. C. M. L. Low, H. R. Drew, M. J. Waring, *Nucleic Acids Res.* **12**, 4865 (1984).
17. K. R. Fox and M. J. Waring, *ibid.*, p. 9271.
18. K. Harshman and P. B. Dervan, *ibid.* **13**, 4825 (1985).
19. P. G. Schultz, J. S. Taylor, P. B. Dervan, *J. Am. Chem. Soc.* **104**, 6861 (1982).
20. J. S. Taylor, P. G. Schultz, P. B. Dervan, *Tetrahedron* **40**, 457 (1984).
21. P. G. Schultz and P. B. Dervan, *J. Biomol. Struc. Dynam.* **1**, 1133 (1984).
22. ———, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6834 (1983).
23. R. S. Youngquist and P. B. Dervan, *ibid.* **82**, 2565 (1985).
24. P. G. Schultz and P. B. Dervan, *J. Am. Chem. Soc.* **105**, 7748 (1983).
25. R. S. Youngquist and P. B. Dervan, *ibid.* **107**, 5528 (1985).
26. E. F. Gale, C. Cundliffe, P-E. Reynolds, M. H. Richmond, M. J. Waring, *The Molecular Basis of Antibiotic Action* (Wiley, New York, 1981), pp. 258–401.
27. S. C. Jain and H. M. Sobell, *J. Mol. Biol.* **296**, 1 (1972).
28. F. Takusagawa, M. Dabrow, S. Neidle, H. M. Berman, *Nature (London)* **296**, 466 (1982).
29. G. Quigley, A. Wang, G. Ughetto, J. van Broom, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7204 (1980).
30. A. H.-J. Wang *et al.*, *Science* **225**, 1115 (1984).
31. G. Ughetto, A. H-J. Wang, G. J. Quigley, G. A. van der Marel, J. H. van Boom, A. Rich, *Nucleic Acids Res.* **13**, 2305 (1985).
32. M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura, R. E. Dickerson, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1376 (1985).
33. P. Kollman, *Acc. Chem. Res.* **18**, 105 (1985).
34. R. P. Hertzberg and P. B. Dervan, *J. Am. Chem. Soc.* **104**, 313 (1982).
35. ———, *Biochemistry* **23**, 3934 (1984).
36. C. Zimmer, in *Progress in Nucleic Acids Research and Molecular Biology*, N. E. Cohn, Ed. (Academic Press, New York, 1975), pp. 285–318.
37. A. K. Krey, in *Progress in Molecular and Subcellular Biology*, F. N. Hahn, Ed. (Springer-Verlag, New York, 1980), vol. 7.
38. J. D. McGhee, *Biopolymers* **15**, 1345 (1976).
39. G. V. Gursky *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **74**, 367 (1982).
40. A. S. Krylov, S. L. Grokhovsky, A. S. Zasedatelev, A. L. Zhuze, G. V. Gursky, B. P. Gottikh, *Nucleic Acids Res.* **6**, 289 (1979).
41. G. Luck, C. Zimmer, K. E. Reinert, F. Arcamone, *ibid.* **4**, 2655 (1977).
42. D. J. Patel and L. L. Canuel, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5207 (1977).
43. L. A. Marky, K. S. Blumenfeld, K. J. Breslauer, *Nucleic Acids Res.* **11**, 2857 (1983).
44. K. Zakrzewska, R. Lavery, B. Pullman, *ibid.*, p. 8825.
45. F. Arcamone, V. Nicolella, S. Penco, S. Redaelli, *Gazz. Chim. Ital.* **99**, 632 (1969).
46. C. Zimmer, G. Luck, E. Birch-Hirschfeld, R. Weiss, F. Arcamone, W. Guschlbauer, *Biochim. Biophys. Acta* **741**, 15 (1983).
47. A. A. Khorlin *et al.*, *FEBS Lett.* **118**, 311 (1980).
48. M. Waring, *J. Mol. Biol.* **54**, 247 (1970).
49. R. M. Wartell, J. E. Larson, R. D. Weiss, *J. Biol. Chem.* **250**, 2698 (1975).
50. D. J. Patel, A. Pardie, K. Itakura, *Science* **216**, 581 (1982).
51. M. A. Krivtsora, E. B. Moroshkina, E. N. Glibin, E. V. Frisman, *Mol. Biol.* **18**, 950 (1984).
52. J. Sluka and P. B. Dervan, *Proceedings of the International Kyoto Conference on Organic Chemistry* (Elsevier, Amsterdam, in press).
53. M. V. Mikhailov, A. S. Zasedatelev, A. S. Krylov, G. V. Gurskii, *Mol. Biol. (Engl. Trans.)* **15**, 541 (1981).
54. R. F. Martin and N. Holmes, *Nature (London)* **302**, 452 (1983).
55. M. K. Dhaon and R. K. Olson, *J. Org. Chem.* **46**, 3436 (1981).
56. M. A. Visiwametra *et al.*, *Nature (London)* **289**, 817 (1981).
57. C. M. L. Low, R. K. Olsen, M. J. Waring, *FEBS Lett.* **176**, 414 (1984).
58. M. J. Waring and L. P. G. Wakelin, *Nature (London)* **252**, 653 (1974).
59. D. H. Ohlendorf, W. F. Anderson, R. G. Fisher, Y. Takeda, B. W. Matthews, *ibid.* **298**, 718 (1982).
60. M. Lewis, A. Jeffrey, J. Wang, R. Ladner, M. Ptashne, C. O. Pabo, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 435 (1983).
61. I. T. Weber and T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3973 (1984).
62. C. A. Frederick *et al.*, *Nature (London)* **309**, 327 (1984).
63. J. E. Anderson, M. Ptashne, S. C. Harrison, *ibid.* **316**, 596 (1985).
64. R. W. Schevitz, Z. Otwinowski, A. Joachimiak, C. L. Lawson, P. B. Sigler, *ibid.* **317**, 782 (1985).
65. N. C. Seeman, J. M. Rosenberg, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 804 (1976).
66. Supported by NIH grant GM-27681, American Cancer Society grant NP-428, and the Burroughs Wellcome Company.