Sigman, D. S., Spassky, A., Rimsky, S., & Buc, H. (1985) *Biopolymers 24*, 183-197.
Spassky, A., & Sigman, D. S. (1985) *Biochemistry 24*, 8050-8056.
Wu, J. C., Stubbe, J. A., & Kozarich, J. W. (1985a) *Biochemistry 24*, 7562-7568.
Wu, J. C., Stubbe, J. A., & Kozarich, J. W. (1985b) *Biochemistry 24*, 7569-7573.

Molecular Recognition between Oligopeptides and Nucleic Acids: Novel Imidazole-Containing Oligopeptides Related to Netropsin That Exhibit Altered DNA Sequence Specificity†

J. William Lown,* Krzysztof Krowicki, and U. Ganapathi Bhat
Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2G2
Andrew Skorobogaty
Department of Biochemistry, La Trobe University, Bundoora, Victoria, Australia 3083
Brian Ward and James C. Dabrowiak
Department of Chemistry, Syracuse University, Syracuse, New York 13120
Received March 27, 1986; Revised Manuscript Received July 29, 1986

ABSTRACT: Oligopeptides have been synthesized that are structurally related to the antiviral antitumor antibiotic netropsin, but in which each of the pyrrole units is successively replaced by an imidazole moiety, as well as their di- and triimidazole-containing counterparts. These compounds bind to duplex DNA with constants in the range (1.06–1.98) × 10^6 M^{-1} but not to single-stranded DNA. Since they bind to T4 DNA, it is inferred that, like the parent antibiotic netropsin, they are also minor groove selective. This series of compounds exhibits a progressively decreasing preference for AT sites in binding studies with both native DNAs and synthetic oligonucleotides and a corresponding increasing acceptance of GC base pairs. Footprinting experiments utilizing a 139 base pair *Hin*dIII/*Nci*I restriction fragment from pBR 322 DNA revealed that these lexitropsins, or information-reading oligopeptides, recognize more sites than the parent netropsin. In addition, some regions of enhanced nuclease action as the result of drug binding to the fragment were identified. The diimidazole compound in particular recognizes GC-rich sites, implying the formation of new hydrogen bonds between G-C(2)NH_2 in the minor groove and the additional N_3 imidazole nitrogens. It is clear however that, since the lexitropsins appear to tolerate the original $(AT)_4$ site, an *N*-methylimidazole group on the ligand will permit either a GC or AT base pair in the binding sequence. Another factor that may be significant in molecular recognition is the high negative electrostatic potential of A·T regions of the minor groove, which is likely to strongly influence binding of these cationic species to DNA. This approach may ultimately permit the structurally rational alteration of sequence specificity in the molecular recognition of oligopeptides for DNA.

The mechanisms whereby peptides, small proteins, and other molecules recognize nucleic acids are fundamental to many important processes in biology and appear to underlie their characteristic properties in living systems (Caruthers, 1980; Ofengand, 1979; Kim et al., 1974; Gursky et al., 1977; Takeda et al., 1983; Frederick et al., 1984). This applies, for example, to DNA–histone interactions (Kim et al., 1974; Gursky et al., 1977), the recognition between enzymes or regulatory proteins and complementary DNA binding sites (Caruthers, 1980; Ofengand, 1979; Kim et al., 1974), and several cases of current clinical interest in anticancer (Neidle & Waring, 1983) and antiviral (Gale et al., 1981) chemotherapy. The binding of such antineoplastic agents as doxorubicin, daunorubicin (Neidle & Waring, 1983; Gale et al., 1981), bleomycin (Hecht, 1979), or neocarzinostatin (Kappen & Goldbert, 1983) to cell target DNA invokes intimate, precise, and highly specific interactions in molecular recognition.

There is increasing knowledge from the molecular biology of gene expression and control, DNA structure and topology, and the recognition of sequences of unusual susceptibility or sensitivity to xenobiotics (Ruddon, 1981). This new information combined with a greater appreciation of structure–activity relationships of anticancer drugs (Neidle & Waring, 1983; Gale et al., 1981) increases the prospects for rational anticancer drug design. Thus, significant advantages would accrue in, for example, the understanding of the processes of gene expression, or the design of vectors for drug targeting, if one could decipher the code controlling the reading of information on DNA by drug molecules. In this regard, a specific goal of ours is to design binding groups to act as vectors for our hemin-based functional bleomycin models (Lown et

†This investigation was supported by grants (to J.W.L.) from the National Cancer Institute of Canada and the Natural Sciences and Engineering Council of Canada and Grant GM 31895 (to J.C.D.) from the National Institutes of Health.

0006-2960/86/0425-7408$01.50/0

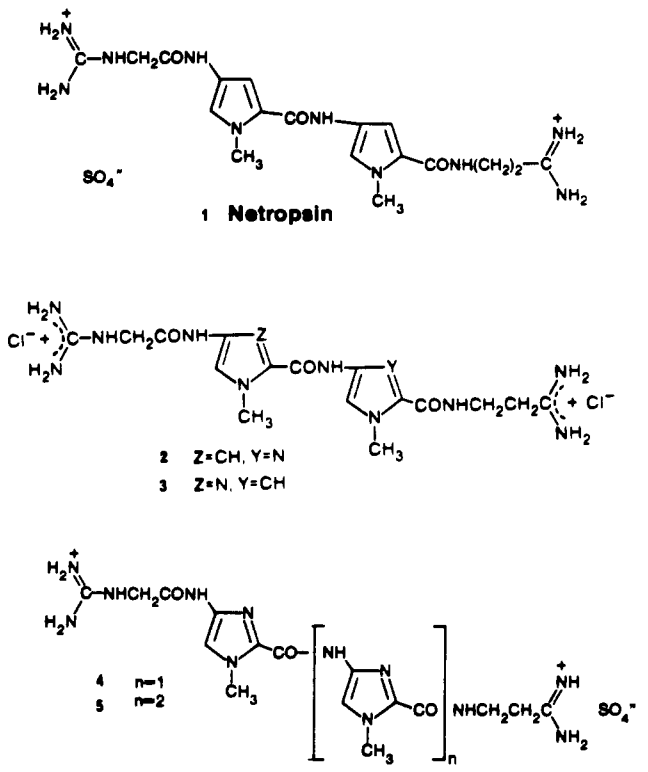

FIGURE 1: Structure of oligopeptide antitumor antibiotic netropsin (**1**) and synthetic oligopeptides **2–5**.

al., 1982, 1984, 1986a) for DNA sites unusually sensitive to lethal double-strand breaks.

The natural oligopeptide antibiotics netropsin (Hahn, 1975) (Figure 1) and distamycin (Hahn, 1979; Arcamone et al., 1967) exhibit antitumor, antiviral, and antibacterial activity. Evidence from a study of their biochemical pharmacology indicates that they act to block the template function of DNA by binding to specific nucleotide sequences in the minor groove of double-helical DNA (Wartell et al., 1974; Zimmer, 1975; Kolchinski et al., 1975; Patel, 1982; Reinert & Thorson, 1970; Kopka et al., 1985; Zimmer et al., 1979b). Since their binding does not involve intercalation, with its concomitant helix distortion (Neidle & Waring, 1983), they offer several advantages as probes of molecular recognition over, for example, the anthracyclines, actinomycins, or antibiotics of the aureolic acid group (Gause, 1975).

Netropsin and distamycin bind within the minor groove of DNA (Hahn, 1975) and demand binding sites consisting of $(AT)_4$ and $(AT)_5$, respectively (Hahn, 1975; Zimmer et al., 1979). An analysis of the structural requirements for the molecular recognition suggested that replacement of one or more pyrrole groups by imidazole should alter this preference for $(AT)_4$ base preference for recognition in a predictable fashion. Accordingly, we report the properties of mono-, di-, and triimidazole-containing oligopeptides related to netropsin, which are in fact observed to exhibit a progressively decreasing preference for AT-rich sites in minor groove binding and, from "footprinting" evidence, also permit recognition of GC-rich sequences. A discussion follows of the factors controlling sequence recognition in these novel oligopeptide vectors.

Materials and Methods

Chemicals. The required oligopeptides **1–5** were synthesized by methods based essentially on our recent syntheses of netropsin, distamycin, and related oligopeptides (Lown & Krowicki, 1985; Lown et al., 1986b). The strategy adopted involves significant changes in the methods and order of introduction of the amidine (Pinner, 1977) and guanidinoacetyl moieties from those reported hitherto (Julia & Preau-Joseph, 1963). This resulted in better yields of the final products. An additional important advantage was the avoidance of chromatography, which is not suitable for such polar compounds because of the likelihood of contamination of the final products with inorganic salts eluted from the absorbents (Patai, 1975).

Melting points were determined on a Fisher-Johns apparatus and are uncorrected. The IR[1] spectra were recorded on a Nicolet 7199 FF spectrophotometer, and only the principal peaks are reported. The ^1H NMR spectra were recorded on Bruker WH-200 and WH-400 cryospectrometers. FAB (fast atom bombardment) mass spectra (Hogg, 1983) were determined on Associated Electrical Industries (AEI) MS-9 and MS-50 double-focusing high-resolution mass spectrometers. Kieselgel 60 (230–400 mesh) of E. Merck was used for flash chromatography, and precoated sheets of silica gel 60F-254 of E. Merck were used for TLC. The TLC system for (i) covalent peptidic compounds was chloroform–methanol (9:1), (ii) ionic compounds with one ionic pair was methanol with some AcOH, and (iii) ionic compounds with two ionic pairs was methanol with some formic acid.

[[1-Methyl-4-[[1-methyl-4-[(guanidinoacetyl)amino]-pyrrol-2-yl]carboxamido]imidazol-2-yl]carboxamido]-propionamidine Hydrochloride (**2**). A solution of 225 mg (0.5 mmol) of [[1-methyl-4-[[1-methyl-4-[(guanidinoacetyl)-amino]pyrrol-2-yl]carboxamido]imidazol-2-yl]carboxamido]propionitrile hydrochloride (Lown et al., 1986b) in ethanol was treated successively with HCl and NH_3 as described earlier. During this procedure, the hydrochloride of the imidoester precipitated from ethanolic solution. The product (192 mg) was dissolved in water (3 mL), and addition of an equivalent of Na_2SO_4 precipitated the product as the sulfate salt. This was recrystallized from water to give 209 mg (79%) of the pure sulfate: mp 230–231 °C; IR (Nujol) 1376, 1404, 1453, 1538, 1663, 3318 (vb) cm^{-1}; MS–FAB, m/z 530 $[(MH)^+]$, 432 $(M-HSO_4)$. A total of 58.1 mg (0.11 mmol) of the sulfate was dissolved in hot water, and 24.4 mg (0.1 mmol) of $BaCl_2 \cdot 2H_2O$ was added. The $BaSO_4$ was collected, and the filtrate was evaporated to dryness. The residue was dissolved in methanol (5 mL) and filtered, the methanol was then removed, and the solid was dissolved in 2 mL of water. Addition of acetone (4 mL) precipitated 47 mg (93% on the basis of $BaCl_2 \cdot H_2O$ used) of pure hydrochloride **2**: mp 186 °C; ^1H NMR ($[^2H_6]Me_2SO$) δ 2.65 (t, 2 H), 3.60 (q, 2 H), 3.86 (s, 3 H), 3.96 (s, 3 H), 4.06 (br s, 2 H), 7.03 (d, 1 H), 7.31 (d, 1 H), 7.46 (br s, ~4 H), 7.54 (s, 1 H), 7.77 (br s, 1 H), 8.14 (t, 1 H), 8.96 and 9.12 (2 br s, 4 H), 10.38 (2 s, 2 H); IR (Nujol) 1377, 1463, 1539, 1610, 1661, 3258 (vb) cm^{-1}; MS–FAB, m/z 432 $[(M - HCl - Cl)^+]$. Anal. Calcd for $C_{17}H_{27}N_{11}O_3Cl_2$: C, 40.5; H, 5.4; N, 3.05; Cl, 14.1. Found: C, 40.1; H, 5.7; N, 30.2; Cl, 13.8.

[[1-Methyl-4-[[1-methyl-4-[(guanidinoacetyl)amino]-imidazol-2-yl]carboxamido]pyrrol-2-yl]carboxamido]-propionamidine Hydrochloride (**3**). The compound [[1-methyl-4-[[1-methyl-4-[(guanidinoacetyl)amino)imidazol-2-yl]carboxamido]pyrrol-2-yl]carboxamido]propionitrile hydrochloride (Lown et al., 1986b) (400 mg, 0.89 mmol) was produced with HCl/EtOH and NH_3/EtOH as in the preceding example, affording 371 mg of a crude product. This

[1] Abbreviations: NMR, nuclear magnetic resonance; MS, mass spectrometry; FAB, fast atomic bombardment; TLC, thin-layer chromatography; IR, infrared; FT, Fourier transform; DMA, dimethylacetamide; Me_2SO, dimethyl sulfoxide; EDTA, ethylenediaminetetraacetic acid; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride.

compound does not form a water-insoluble sulfate. However, it can be prepared as the sulfate salt by precipitation from a 10% methanolic solution with 1 equiv of tetraethylammonium sulfate.[2] The crude product (371 mg) was dissolved in MeOH (5 mL), and a few drops of water were added (to prevent the formation of a salt, which is very hygroscopic and dissolves if exposed to atmospheric moisture almost immediately). Upon the addition of a methanolic solution of tetraacetylammonium sulfate, a jelly-like substance formed. This was collected by filtration, washed with methanol and ethyl acetate, and dried under vacuum at 100 °C to give 330 mg (70% yield) of the product in the form of the sulfate: no distinct mp (it starts to decompose at 235 °C); IR (Nujol) 1122, 1376, 1461, 1547, 1577, 1672, 3250 (vb) cm^{-1}; MS–FAB, m/z 530 $[(MH)^+]$, 432 $[(M - HSO_4)^+]$.

The sulfate salt of the product (35 mg, 0.66 mmol) was dissolved in water, and $BaCl_2 \cdot 2H_2O$ (14.6 mg, 0.06 mmol) was added. The barium sulfate formed was collected, and the filtrate was evaporated to dryness. The product was extracted with methanol, the extracts were evaporated, and the residual solid was extracted with absolute EtOH. The ethanol was removed, and the residue was dissolved in a drop of water. Upon addition of acetone, the pure product was collected and dried under vacuum at 100 °C to give **3**: 24 mg (79% yield); mp 184–185 °C; ^1H NMR ($[^2H_6]Me_2SO$) δ 2.70 (t, 2 H), 3.54 (q, 2 H), 3.84 (s, 3 H), 3.97 (s, 2 H), 4.11 (s, 2 H), 7.07 (d, 1 H), 7.26 (d, 1 H), 7.35 (br s) and 7.46 (s) (together 5 H), 7.76 (br s, 1 H), 8.32 (t, 1 H), 8.78 and 9.07 (2 br s, 4 H), 9.88 (s, 1 H), 10.53 (s, 1 H); IR (Nujol) 1376, 1462, 1559, 1580, 1658, 1679, 3140, 3250, 3320 cm^{-1}; MS–FAB, m/z 432 $[(M - HCl - Cl)^+]$. Anal. Calcd for $C_{17}H_{27}Cl_2N_{11}O_3$: C, 40.5; H, 5.4; N, 30.5; Cl, 14.1. Found: C, 40.2; H, 5.7; N, 30.1; Cl, 13.7.

[[1-Methyl-4-[[1-methyl-4-[(guanidinoacetyl)amino]imidazol-2-yl]carboxamido]imidazol-2-yl]carboxamido]propionamidine Sulfate (**4**). A solution of 650 mg (1.44 mmol) of [[1-methyl-4-[[1-methyl-4-[(guanidinoacetyl)amino]imidazol-2-yl]carboxamido]imidazol-2-yl]carboxamido]propionitrile hydrochloride (Lown et al., 1986b) in 25 mL of anhydrous ethanol was saturated with dry HCl gas while being cooled efficiently. After 1.5 h at room temperature, the solvent was removed under reduced pressure, and the residual solid was washed with dry ether and then decanted and redissolved in 20 mL of dry ethanol. Dry ammonia gas was condensed into the vessel; after 1 h at room temperature, the solution was evaporated to dryness, and the residual solid crystallized when dissolved in a large volume of methanol and was condensed to a small volume to afford 650 mg of the hydrochloride derivative of **4** (**4A**): mp 190–193 °C; ^1H NMR ($[^2H_6]Me_2SO$) δ 2.70 (t, 2 H), 3.58 (q, 2 H), 3.96 and 3.98 (2 s, 6 H), 4.17 (d, 2 H), 7.51 (br m, 6 H), 7.90 (t, 1 H), 8.54 (t, 1 H), 8.92 and 9.15 (2 br s, 4 H, amidine), 9.34 (s, 1 H), 10.90 (s, 1 H) (D_2O exchange reveals 2 s at 7.52 and 7.56); MS–FAB, m/z 433 $[(M - HCl - Cl)^+]$.

The hydrochloride salt **4A** was dissolved in water, and a solution of sodium sulfate was added to precipitate the crystalline **4**. The latter was purified by recrystallization from 80 mL of water to give pure **4**: 650 mg (73.7% yield); mp 250 °C dec; MS–FAB (glycerol), m/z 531 $[(MH)^+]$, 433 $[(M - HSO_4)^+]$. Anal. Calcd for $C_{16}H_{26}N_{12}O_7S$: C, 36.2; H, 4.9; N, 31.7; S, 6.0. Found: C, 35.9; H, 5.1; N, 31.6; S, 6.2.

[[1-Methyl-4-[[1-methyl-4-[[1-methyl-4-[(guanidinoacetyl)amino]imidazol-2-yl]carboxamido]imidazol-2-yl]carboxamido]imidazol-2-yl]carboxamido]propionamide Sulfate (**5**). The compound [[1-methyl-4-[[1-methyl-4-[[1-methyl-4-nitroimidazol-2-yl]carboxamido]imidazol-2-yl]carboxamido]imidazol-2-yl]carboxamido]propionitrile (Lown et al., 1986b) (270 mg) was reduced with 3.5 g of $SnCl_2 \cdot 2H_2O$ in 6.3 mL of HCl at 5 °C for 15 min. With efficient cooling 25 mL of 25% aqueous NaOH was added, and the mixture was extracted with methylene chloride. The extracts were evaporated to a small volume, and addition of hexane precipitated 202 mg of the amine (80% yield). The amine at this stage was contaminated with a little of the chloro derivative as revealed by TLC on silica gel with 10% MeOH in $CHCl_3$ as eluent. The amine (202 mg, 0.46 mmol) was dissolved in 2.5 mL of dimethylacetamide (DMA), and 141 mg (0.92 mmol) of guanidinoacetic acid hydrochloride was added. Then a solution of 141 mg (0.68 mmol) of dicyclohexylcarbodiimide (DCC) in 1 mL of DMA was added to the mixture during 2 h, and after being stirred for an additional 1 h, the solvents were removed in vacuo. The residue was dissolved in water, the dicyclohexylurea removed by centrifugation, and the aqueous layer evaporated to dryness. The residue was dissolved in hot methanol, decolorized with charcoal, and on cooling gave 112 mg of solid. Recrystallization from methanol gave [[1-methyl-4-[[1-methyl-4-[[1-methyl-4-[(guanidinoacetyl)-amino]imidazol-2-yl]carboxamido]imidazol-2-yl]carboxamido]imidazol-2-yl]carboxamido]propionitrile hydrochloride: 90 mg (34% yield); ^1H NMR ($[^2H_6]Me_2SO$) δ 2.78 (t, 2 H, J = 6 Hz), 3.49 (q, 2 H, J = 6 Hz), 3.97, 3.99 and 4.02 (3 s, 9 H), 4.07 (br s, 2 H), 7.28 (br s, 4 H), 7.53, 7.56, 7.66 (3 s overlapped with a br s, 4 H), 8.55 (t, 1 H, J = 6 Hz), 9.64 and 9.68 (2 s, 2 H), 10.70 (s 1 H); MS–FAB, m/z 539 $[(M - Cl)^+]$.

A solution of 90 mg (0.16 mmol) of the base in 5 mL of dry ethanol was saturated with dry HCl at <0 °C. After 1 h at room temperature, the solvent was removed under reduced pressure, the residue was washed with dry ether and dissolved in 5 mL of dry ethanol, and dry NH_3 gas was condensed into the vessel. After 1 h at room temperature, the solvents were evaporated, and the residue was recrystallized from methanol. This solid was taken up in 2 mL of water, and 2 mL of 20% sodium sulfate solution was added to precipitate a gel, which was collected by centrifugation. The gel was redissolved in 10 mL of water and concentrated to a small volume. Then, when heated, a white precipitate formed (probably due to a different degree of solvation). The solid was collected, washed with water and then ethanol, and dried to give **5**: 53 mg (50% yield); mp 265 °C dec; MS–FAB, m/z (relative intensity) 654 (2.7, MH^+) 556 [16.8, $(M - HSO_4)^+$]. The ^1H NMR could not be obtained because of the low solubility in all suitable solvents. Anal. Calcd for $C_{21}H_{31}N_{15}O_8S$: C, 38.6; H, 4.8; N, 32.1; S, 4.90. Found: C, 38.5; H, 4.9; N, 31.9; S, 4.9.

Biochemicals. The natural DNAs, *Micrococcus luteus Escherichia coli*, *Clostridium perfringens*, T4, and calf thymus DNAs as well as the synthetic polynucleotides were obtained from Sigma. These were dialyzed twice against appropriate buffers at 5 °C and lyophilized before use. The plasmid DNA pBR 322 and calf thymus DNA (type 1, Na^+ salt) were obtained from Bethesda Research Laboratories and Sigma, respectively. The enzymes *Ncl*, *Hin*dIII, and AMV reverse transcriptase were purchased from Bethesda Research Laboratories. The radiolabeled nucleotide [α-^{32}P]dATP was purchased from New England Nuclear.

[2] Preparation of tetraethylammonium sulfate: tetraethylammonium hydroxide (5 mL, 20% aqueous solution) and sulfuric acid (190 μL of H_2SO_4 in 1 mL of water) were mixed together, and the water was evaporated to give a thick semicrystalline syrup.

Table I: ΔT_m (deg) Values for Oligopeptides Bound to Native DNAs and Synthetic Polynucleotides[a]

compd	native DNA (% G+C) (D/P = 0.25:1)			
	C. perfringens (26.5)	calf thymus (42)	*E. coli* (50)	*M. luteus* (72)
netropsin	28	21.5	17.5	7
2	24.5	18.5	16	7.5
3	24	18	16	7.5
4	18	15.5	13.5	8
5	15	12	12.5	8.5

	polydeoxyribonucleotide										
	D/P for poly(dA)·poly(dT)[d]			D/P for poly(dA-dT)·poly(dA-dT)			D/P for poly(dG)·poly(dC)			D/P for poly(dG-dC)·poly(dG-dC)	
compd[b]	0.5:1	0.25:1	0.1:1	0.5:1	0.25:1	0.1:1	0.5:1	0.25:1	0.1:1	0.1:1	0.05:1
netropsin	[c]	61.5	55	45.5	42	37.5	8	6	3	9	5
2	58.5	54	47.5	38	35	31	10	8	5	14	8
3	55.5	51.5	45.5	39	36	32	11.5	9.5	6	16	9
4	46	41	34	28.5	25.5	20.5	12	12	7	16	9

[a] DNA, ca. 1.0 A_{260}, samples (0.3 mL) in 40 mM potassium–sodium phosphate buffer and Na_2EDTA, pH 6.9 (except where indicated), purged with helium. D/P = drug to phosphate residue ratio. [b] Compound **5** caused precipitation with the oligomers, preventing measurement of ΔT_m values. [c] Temperature elevation too high to measure because of instrumental limitations. [d] For a DNA concentration of 1.0 A_{260}, it was necessary to reduce buffer–salt concentration to 5 mM to accommodate elevated melting temperature.

DNA Thermal Denaturation Determinations. Solutions of the dialyzed and lyophilized natural DNAs were prepared by stirring with 40 mM potassium–sodium phosphate buffer, pH 6.9, containing 2 mM Na_2EDTA overnight. The solutions of poly(dA)·poly(dT) and poly(dG-dC)·poly(dG-dC) were prepared in buffers in a similar fashion. Concentrations of the DNAs were determined spectroscopically and kept at ca. 1.0 A^{260}. The concentrations of the buffer used with AT-rich polymers were kept at 5 mM in order to accommodate the elevated T_m values with the lexitropsins.

Thermal denaturation profiles (T_m) were measured at drug:DNA phosphate (D/P) values of 0.1:1, 0.25:1, and 0.5:1 and in some cases 0.05:1 on a Gilford Model 2400 spectrophotometer equipped with a thermostated cell compartment. DNA samples (0.3 mL) in 40 mM sodium–potassium phosphate buffer, pH 7.0, and 50 mM sodium chloride were purged with helium and overlaid with paraffin oil. When the temperature was raised at 0.5–1.0 deg/min, the T_m's were reproducible to ±1 °C.

Determination of Binding Constants to DNA. The relative binding constants were estimated by displacement of intercalative binding of ethidium to calf thymus DNA and employment of a value of $K_{assoc} = 1 \times 10^6$ M^{-1} at pH 7.0, 37 °C, and 40 mM NaCl for ethidium bound to calf thymus DNA (Le Pecq & Paoletti, 1967). It was determined that none of the oligopeptides interferes with the fluorescence measurements, which were performed on a Turner 430 spectrofluorometer. The procedure, which involves following the displacement of the ethidium upon being titrated in the drugs and determining the concentration of drug required to displace 50% of the ethidium, follows that of Morgan et al. (1979) and gives relative rather than absolute values for binding constants. Higher concentrations of lexitropsins displace all the ethidium from the DNA.

DNase I Footprinting. The concentrations of **1**, **3**, **4**, and DNA (in base pairs) were determined optically by using extinction coefficient values of (in mM^{-1} cm^{-1}) $\epsilon_{296} = 20.2$, $\epsilon_{299} = 14.6$, $\epsilon_{305} = 23.1$, and $\epsilon_{260} = 13.1$, respectively. The *Hin*dIII/*Nci*I, 139 base pair restriction fragment from pBr 322 DNA was isolated and end-labeled in the earlier described manner (Lown et al., 1986a).

Solutions for footprinting experiments were prepared by mixing the following: 2 μL of calf thymus DNA, 770 μM base pairs in 50 mM Tris-HCl pH 7.0 buffer; 2 μM of fragment having a concentration of ~7 μM in buffer; 2 μL of a buffered solution of the compound; 2 μL of activated enzyme complex. The enzyme complex was prepared by addition of 2 μL of a stock solution of the enzyme in 50% glycerol to 398 μL of buffer containing 50 mM Tris-HCl, 32 mM $MgCl_2$, and 8 mM $CaCl_2$, pH 7.0. Prior to addition of enzyme and digestion for 10 min at 37 °C, the DNA–compound mixtures were preincubated for 90 min at 37 °C. For each compound, 10–15 separate digests at different values of r_t were carried out, where r_t is the ratio of compound to DNA base pairs. The digests were terminated by the addition of 10 μL of a solution containing 70% aqueous urea, 20 mM EDTA, and 0.025% each of bromphenol blue and xylene cyanol followed by heating to 80 °C for 10 min. The extent of reaction was such that ~80% of the labeled strand of the restriction fragment remained uncleaved during the digestion (Lane et al., 1983; Goodisman & Dabrowiak, 1985).

Electrophoresis of the digested radiolabeled 139-mer using an in-house developed thermostated field gradient electrophoresis device was as previously described (Ward et al., 1986). Following autoradiography, the negative was analyzed with a linear scanning microdensitometer (Dabrowiak et al., 1986) to yield scans showing the relative concentrations of the various oligomers produced in the digests. Establishment of sequence involved guanine-specific cleavage of DNA (Maxam & Gilbert, 1980) and reference to the cleavage pattern of the fragment in the presence of DNase I (Lown et al., 1986a). Owing to poor resolution of the longer labeled oligomers in the gel, cleavages at position >99 of the fragment were not analyzed.

Results

Minor Groove and Base Preferential Binding of Novel Oligopeptides to Native DNAs and Synthetic Oligodeoxyribonucleotides. The oligopeptides **2**–**5** (Figure 1) bind to duplex DNA but not to single-stranded DNA. Appreciable elevations in thermal denaturation temperatures (ΔT_m) are obtained with the oligopeptides upon interaction with T4 DNA. Netropsin and compounds **2**–**5** give the following ΔT_m readings with T4 DNA at a D/P (drug/phosphate residue) ratio of 0.25:1, respectively: 25, 23, 21, 18, and 14 deg. Since T4 DNA has glycosylation of the 5-(hydroxymethyl)cytidine residues, which occlude the major groove (Erickson & Szybalski), the observed binding is consistent with attachment in

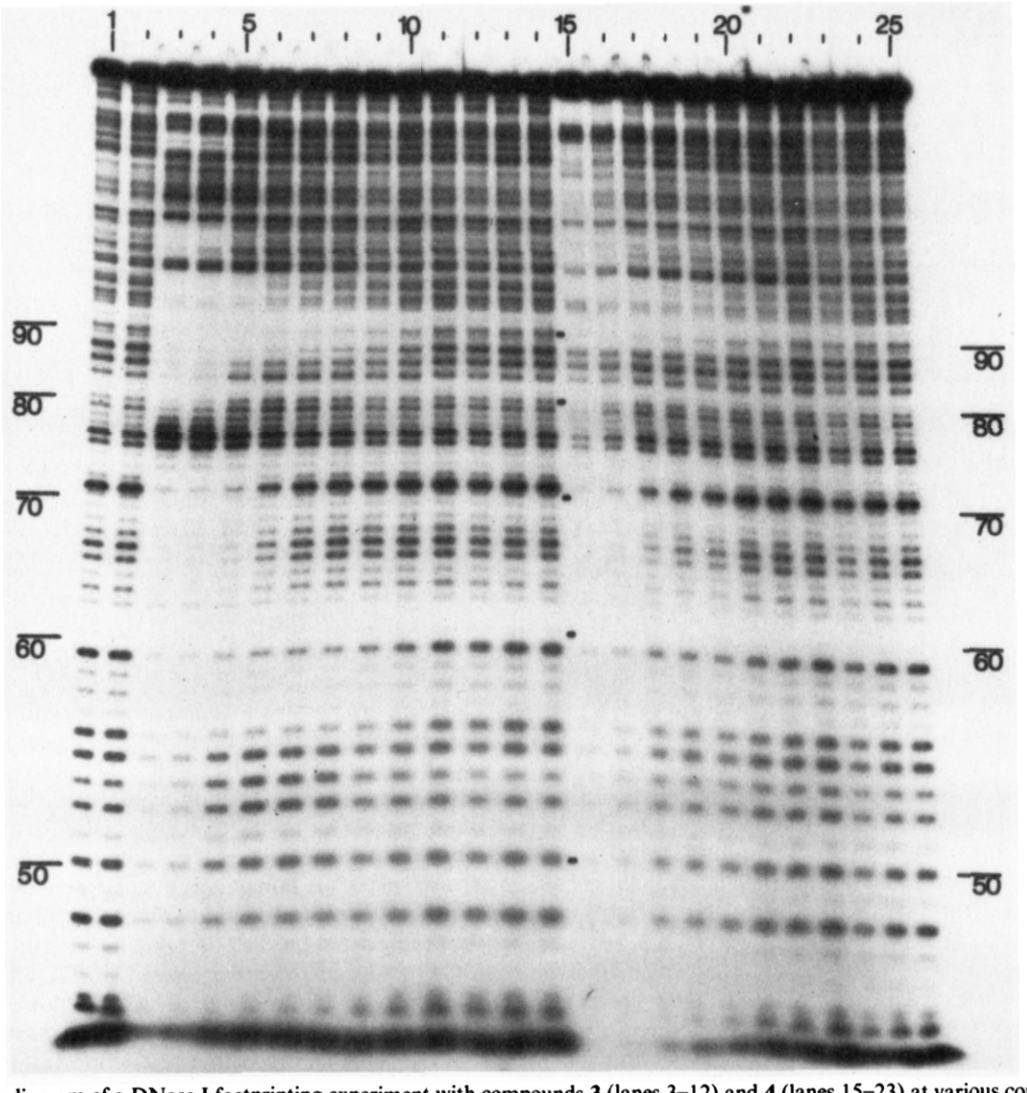

FIGURE 2: Autoradiogram of a DNase I footprinting experiment with compounds **3** (lanes 3–12) and **4** (lanes 15–23) at various compound-to-DNA base pair ratios (r_t) is shown. (Lanes 1, 2, 13, 14, 24, and 25) DNase I digests of the 139 base pair restriction fragment from pBR 322 DNA in the absence of compound. The lane numbers and values of r_t (in parentheses) for the various experiments were as follows: 3 (0.88); 4 and 15 (0.5); 5 and 16 (0.25); 6 and 17 (0.17); 7 and 18 (0.13); 8 and 19 (0.10); 9 and 20 (0.067); 10 and 21 (0.05); 11 and 22 (0.02); 12 and 23 (0.01).

the minor groove as has been established for netropsin (Kopka et al., 1985; Zimmer et al., 1979).

Effects of Oligopeptides on T_m of Native DNAs of Different Base Composition at pH 6.9 and at a D/P Ratio of 0.25. The data in the Table I show a progressive lessened preference for binding of the oligopeptides for A·T sites in the native and synthetic DNAs as each of the pyrrole moieties is successively replaced by imidazole. The strongest trend toward lessened preference for A·T sites (and a corresponding G·C acceptance) is observed for the triimidazole **5**. Although it is recognized to be difficult to estimate binding constants from T_m data (McGhee, 1976; Lee & Waring, 1978), the substantial ΔT_m values at a D/P ratio of 0.25 are in accord with binding comparable with that of the parent netropsin by all the compounds. The ΔT_m values observed with T4 DNA at 25, 23, 21, 18, and 14 deg for netropsin, **2**, **3**, **4**, and **5**, respectively, may be compared with the values of 25, 22, 17, and 13 deg predicted from the table on the basis of the 34% G+C content of T4 from the results obtained with the native DNAs.

Table I also shows the effects of binding of the oligopeptides on the T_m of synthetic repeating sequence DNAs, again signifying a decreased preference for A·T sites. The relative binding constants of the oligopeptides to calf thymus DNA were determined by displacement of intercalative binding of ethidium and have the following values: netropsin, 1.87×10^6 M^{-1}; **2**, 1.87×10^6 M^{-1}; **3**, 1.98×10^6 M^{-1}; **4**, 1.76×10^6 M^{-1}; **5**, 1.06×10^6 M^{-1}. The binding, while strong for all the compounds, decreases somewhat upon successive introduction of imidazole units, in contrast to the effects of adding pyrrole moieties in the case of the parent antibiotic (Arcamone et al., 1967).

Binding Specificity of Oligopeptides Studied Using DNase I Footprinting Methodology. Figure 2 shows an autoradiogram of footprinting experiments involving the netropsin analogues **3** and **4** and the 139 base pair restriction fragment of pBR 322 DNA. Densitometric scans of the autoradiographic data for compounds **1**, **3**, and **4** are presented In Figure 3–5, and the results of the footprinting studies are summarized on the sequence of the fragment, Figure 6.

Inspection of the densitometric data for netropsin revealed that at low values of the drug-to-DNA base pair ratio, r_t, $0 < r_t < \sim 0.04$, inhibitions at positions (all 5′ → 3′) 50–45 (ATAAAC), 62–56 (AATTTAA) and 94–85 (AGATTTCATA) were observed. As r_t was increased, $\sim 0.04 < r_t < \sim 0.17$, no further changes in the DNase I digest pattern of the fragment were observed. As is evident from Figures 2 and 6, netropsin also induced increases in cleavage at positions 78–64, having the sequence GCCTGACTGCGTTAG.

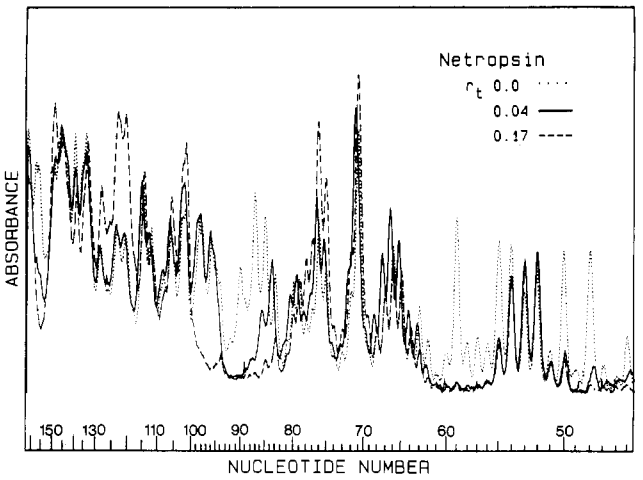

FIGURE 3: Densitometric scans of footprinting experiments involving netropsin (**1**) and 139 base pair restriction fragment of PBR-322 DNA are shown. The numbering system is the genomic numbering system of pBR 322 DNA (Maniatis et al., 1982).

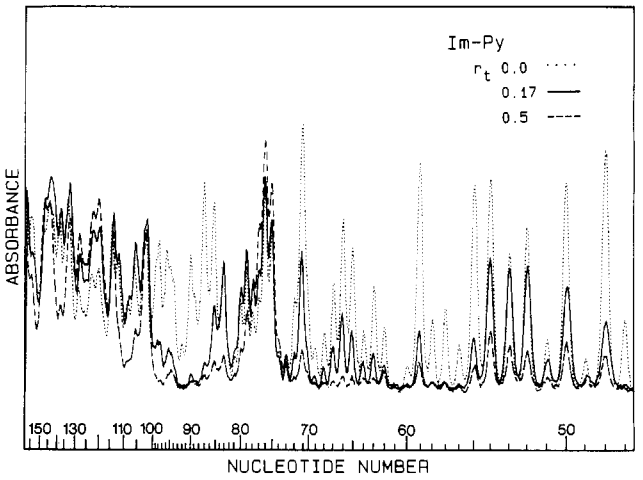

FIGURE 4: Densitometric scans of footprinting experiments involving compound **3** and the 139 base pair restriction fragment of pBR 322 DNA are shown.

Inspection of the densitometric data revealed that inhibitions for compounds **3** and **4** generally occurred at values of r_t that were less than those of netropsin. Thus, for r_t values in the range $0 < r_t < \sim 0.17$, compounds **3** and **4** inhibited the enzyme in the same regions as netropsin but, in addition, exhibited binding sites at ~70 and at 90–95, Figure 6. As compound was added, $\sim 0.17 < r_t < \sim 0.5$, additional inhibitions in GC-rich regions were observed for both analogues. Noteworthy was the fact that at high values of r_t the bis(*N*-methylimidazole) compound, **4**, inhibited the enzyme at nearly every nucleotide position of the fragment, Figures 5 and 6. The binding of compound **3** to the fragment produced enhanced cleavage at positions 78–74 (GCCTG).

Discussion

The results summarized in Table I show that as each pyrrole unit is successively replaced by an imidazole moiety the resulting oligopeptides show an overall decreased AT preference and an increasing acceptance of GC base pairs in binding to DNA. Since the binding was shown to occur in the minor groove, the 2-NH_2 group of guanine is implicated in this process (Sobel & Jain, 1972). These data reflect average base preferences over all permitted acceptable sites, but more detailed information is obtained from the footprinting experiments.

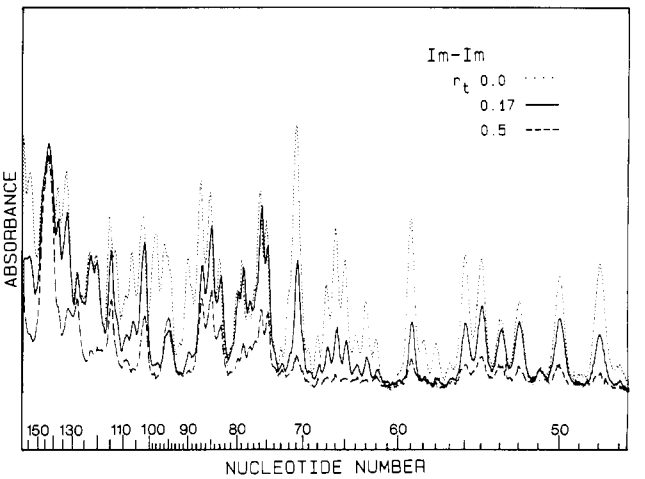

FIGURE 5: Densitometric scans of footprinting experiments involving compound **4** and the 139 base pair restriction fragment of pBR 322 DNA are shown.

Although it was not possible to ascertain from the data presented the nature of the contacts between compounds **3** and **4** and DNA, a few observations pertaining to the interaction are possible. X-ray crystallographic studies of netropsin bound to the dodecanucleotide duplex 5′-CGCGAATTCGCG-3′ reveal that the antibiotic binds within the minor groove by displacing water of hydration (Kopka et al., 1985). The overall reaction may be expressed as $DNA{\cdot}xH_2O + Nt{\cdot}yH_2O \rightarrow DNA{\cdot}Nt + (x + y)H_2O$. X-ray diffraction evidence on a B-DNA dodecamer and on the free netropsin (Berman et al., 1979) suggests values of $x \approx 12$ and $y \approx 5$. The binding mechanism involves two separate processes. Hydrogen bonds from the amide NH groups bridge the strands to the exposed adenine N(3) and thymine O(2) (Kopka et al., 1985) on adjacent sites, thus mimicking the bridging provided by the displaced water of hydration (Figure 7). Additional hydrogen bonding exists between the terminal guanidino and amidino groups to adenine N(3) centers flanking the sequence recognized (Kopka et al., 1985). This striking feature of bidentate hydrogen bonding is also a characteristic of the binding of the amino acid side chains in Cro repressor to its recognition site (Takeda et al., 1983). Such bidentate and multidentate interactions provide a clear rationale for enhancing the specificity of DNA–protein recognition (Helene, 1977; von Hippel, 1979; Day et al., 1973).

The actual sequence recognition or semantophore property accounting for the observed AT sequence preference of the antiviral agent netropsin (Lane et al., 1983; van Dyke et al., 1982; Zimmer, 1975) has, as its structural origin, van der Waals contacts between the C(2) hydrogens of the pyrrole groups and the methylene hydrogens flanking the bis(pyrrole) moiety of the drug and the minor groove of DNA (Figure 7a). Dickerson has suggested (Kopka et al., 1985) that, from the informational standpoint, the role of hydrogen bonding here is not recognition of base sequence but rather proper orientation of the reading frame so that the base sequence information can be read out by van der Waals contacts. Such hydrophobic interactions have also been invoked in the recognition of the *lac* repressor (Wartell et al., 1974).

If guanine is present in the binding sequence, the 2-amino group of the heterocycle, which is significantly above the floor of the minor groove, sterically prevents close approach of the drug to DNA. It is this factor that excludes G from the binding sequence, thus confering AT specificity to the drug. We have proposed that this situation can be altered by including into the framework of a netropsin-type ligand hydrogen

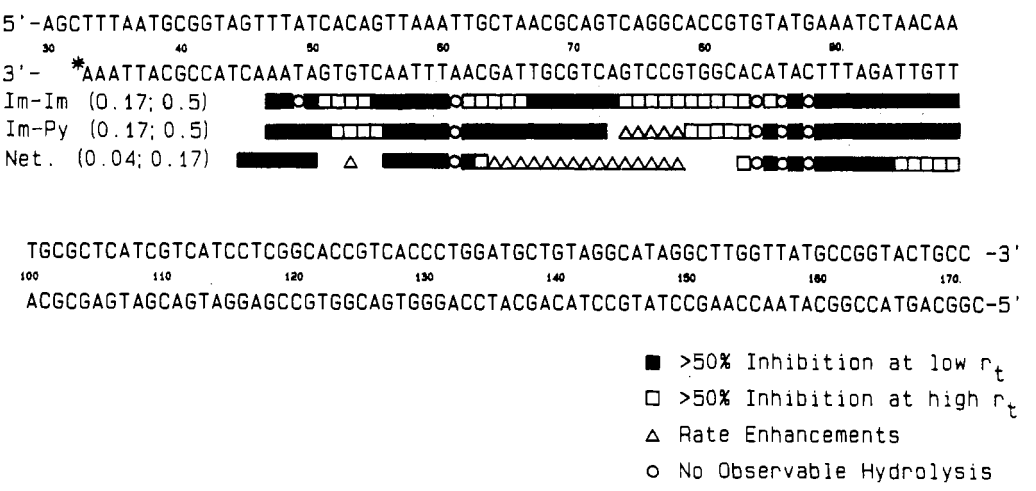

FIGURE 6: Summary of binding information on the sequence of the restriction fragment is shown.

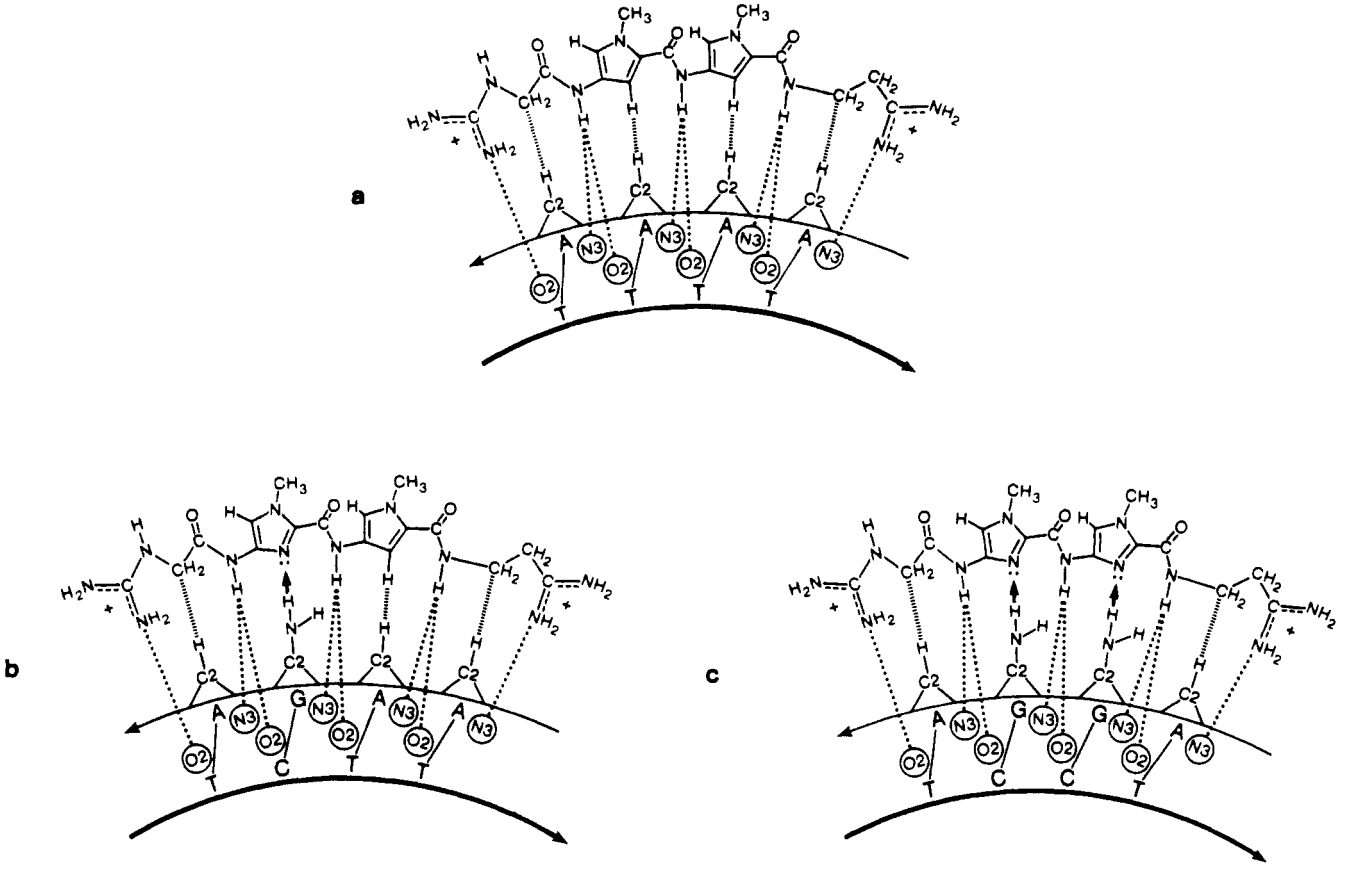

FIGURE 7: Depiction of idealized molecular recognition characteristics and duplex DNA base selectivity for oligopeptides: (a) netropsin bound to the sequence 3′-AAAA-5′, (b) imidazole-substituted **3** bound to the sequence 3′-AGGA-5′. (···) Specific hydrogen bonds; (≡) van der Waals recognition contacts; (→) additional specific G–C(2)NH_2 hydrogen bonds directed from the minor groove.

bond acceptors capable of binding to the 2-amino group of guanine. This is the case for both **3** and **4**, which, due to the presence of the *N*-methylimidazole group(s), should be capable of identifying a guanine, via hydrogen bonding, in the binding sequence. Ideally, **3**, possessing a single *N*-methylimidazole moiety, would bind to a 4 base pair AT-rich sequence containing a single interior GC base pair while **4**, having two *N*-methylimidazole groups, would identify a 4 base pair sequence having *two* interior GC base pairs each flanked by AT base pairs (Figure 7b,c).

Figure 4 shows that compounds **3** and **4** bind to DNA with less specificity than that of netropsin and that both are able to recognize guanine in the binding sequence. Of the two analogues, the bis(*N*-methylimidazole) compound, **4**, is the less specific and at high values of r_t is able to inhibit the enzyme at all positions of the fragment. However, the location of GC base pairs within the binding sequences of compounds **3** and **4** does not appear to support guanine *N*-methylimidazole hydrogen bonding as being the sole basis for the observed altered specificity of the compounds. Consideration of the inhibition regions for **3** and **4** binding sites revealed that, with the possible exceptions of the sites centered at position ~91 and ~96, none of the sites possess sequences predicted by the aforementioned binding analysis. In addition, the existence of the small site centered at position 122 with sequence 3′-CGTGG-5′ for **4** indicates that the idealized contacts shown in Figure 7 may not be correct. Since the single A·T base pair is located in the center of the sequence, it would appear that

the methylene hydrogens flanking the bis(*N*-methylimidazole) moiety of the ligand would enter into steric contact with the guanine amino groups of the G·C base pairs of the binding sequence. On the basis of the structure of the ligand and its conformation on the sequence, it would appear that these contacts would jeopardize binding at the sequence.

Both netropsin and compound **3** give rise to apparent increases in the cleavage rate of DNase I away from the sites of ligand binding. Although further study is required, these intensity enhancements may be the result of DNA structural changes induced by ligand binding and/or alterations of the ratio of the enzyme concentration to the available substrate DNA concentration. Since netropsin, as well as compounds **3** and **4**, would be expected to cause little or no distorttion of DNA (Kopka et al., 1985), the latter is probably the cause of the increased cleavage away from sites of ligand binding.

More detailed analysis of the extent to which the new hydrogen bonds compensate thermodynamically for the loss of the van der Waals contacts between AC(2) and the pyrrole C(3)H must await determination of the spatial dimensions of bonding in the ongoing X-ray diffraction studies of **2** and **3** cocrystallized with a duplex dodecamer in collaboration with Professor Dickerson. Kopka et al., (1985) have independently suggested the substitution of imidazole for pyrrole in netropsin on the basis of X-ray diffraction analysis of the antibiotic cocrystallized with a dodecamer.

Clearly, factors other than those alluded to are important in the binding process. For example, crystallographic and calorimetric studies have shown that ligand and DNA solvation (entropic effects), unaccounted for in the analysis, play an important role in ligand–DNA interactions (Kopka et al., 1985; Marky et al., 1983; Westhof et al., 1985). The requirements of binding of the charged end groups to adenine N(3) centers impose a conflicting base requirement. The gradual change in base preference upon replacing pyrrole by imidazole seen the table suggests that this opposing influence is substantial. An additional observation concerns the fact that the minor groove of A·T-rich regions of DNA possesses a high negative electrostatic potential (Laverly & Pullman, 1985), which is likely to strongly influence binding of cations to DNA and which is in accord with the data of Table I. Such ionic interactions may be utilized by peptidic agents in sliding along the DNA to reach the target site (Ohlendorf et al., 1982; Berg et al., 1981). Thus, in addition to introducing guanine "reading" groups in the ligand, it may be necessary to overcome this A·T bias by reduction of the total cationic charge on the ligand as well. Some structural modifications of this type on the parent antibiotics have been explored. Netropsin with the cationic ends removed and analogues with cationic ends but without hydrogen bonding capabilities both exhibit an appreciable binding to B DNA (Hahn, 1975). A logical extension of this approach to preselected chemical recognition is therefore the replacement of the guanidino and/or amidino moieties by end groups designed spatially to bind to G or C hydrogen bond acceptors. The synthesis of such compounds is now in progress.

Conclusions

The results suggest that it may be possible ultimately to synthesize oligopeptides to bind to a predetermined sequence specifically. Such an important goal will not be achieved easily, however, since the molecular recognition processes are likely to be complex. For example, Cro and λ repressors apparently use a nonidentical set of amino acid side chains organized in a somewhat different spatial arrangement to recognize the same DNA sequence (Anderson et al., 1983; Lewis et al., 1982; Takeda et al., 1983). This suggests that there is not a simple one-on-one recognition code between peptides and bases. The utility of this approach to drug targeting will be reported in due course.

Registry No. **1**, 1438-30-8; **2**, 104394-12-9; **2** (free base), 101772-40-1; **2** (nitrile precursor), 104394-02-7; **2** (sulfate), 101772-41-2; **3**, 101809-75-0; **3** (free base), 104394-04-9; **3** (nitrile precursor), 104394-03-8; **3** (sulfate), 104394-05-0; **4**, 101772-43-4; **4** (free base), 101772-42-3; **4** (nitrile precursor), 104394-06-1; **4A**, 104394-07-2; **5**, 104394-13-0; **5** (free base), 104394-11-8; **5** (nitrile precursor), 104394-08-3; **5** (nitrile precursor amine), 104394-09-4; **5** (guanidinoacetylaminonitrile precursor), 104394-10-7; $NH{=}C(NH_2)NHCH_2CO_2H{\cdot}HCl$, 14901-20-3.

References

Anderson, W. F., Ohlendorf, D. H., Cygler, M., Takeda, Y., & Matthews, B. W. (1983) *UCLA Symp. Mol. Cell. Biol., New Ser. 8*, 19–32.

Arcamone, F., Orezzi, P. G., Barbieri, W., Nicolella, V., & Penco, S. (1967) *Gazz. Chim. Ital. 97*, 1097.

Beman, H. M., Neidle, S., Zimmer, C., & Thrum, H. (1979) *Biochim. Biophys. Acta 561*, 124.

Berg, O. G., Winter, R. B., & von Hippel, P. H. (1981) *Biochemistry 20*, 6929–6947.

Caruthers, M. H. (1980) *Acc. Chem. Res. 13*, 155.

Day, R. O., Seeman, N. C., Rosenberg, J. M., & Rich, A. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 849.

Erickson, R. L., & Szybalski, W. (1964) *Virology 22*, 111.

Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B. C., Greene, P., Boyer, H. W., & Rosenberg, J. M. (1984) *Nature (London) 309*, 327.

Gale, E. F., Cundliffe, E., Reynolds, P. E., Richmond, M. H., & Waring, M. J. (1981) *The Molecular Basis of Antibiotic Action*, 2nd ed., Wiley, New York.

Gause, G. F. (1975) in *Antibiotics III. Mechanism of Action of Antimicrobial and Antitumor Agents* (Corcoran, J. W., & Hahn, F. E., Eds.) p 197, Springer-Verlag, New York.

Goodisman, J., & Dabrowiak, J. C. (1985) *J. Biomol. Struct. Dyn. 2*, 967–979.

Gurskii, G. V., Tumanyan, V. G., Zasedatelev, A. S., Zhuze, A. L., Grokhovsky, S. L., & Gottikh, B. P. (1977) in *Nucleic Acid-Protein Recognition* (Vogel, H. J., Ed.) p 189, Academic, New York.

Hahn, F. E. (1975) in *Antibiotics III. Mechanism of Action of Antimicrobial and Antitumor Agents* (Corcoran, J. W., & Hahn, F. E., Eds.) pp 79–100, Springer-Verlag, New York.

Hecht, S. M., Ed. (1979) *Bleomycin, Chemical, Biochemical and Biological Aspects*, Springer-Verlag, New York.

Helene, C. (1977) *FEBS Lett. 74*, 10.

Hogg, A. M. (1983) *Int. J. Mass Spectrom. Ion Phys. 49*, 25.

Julia, M., & Prêau-Joseph, N. (1963) *C. R. Acad. Sci., Ser. 3 257*, 1115.

Kappen, L. S., & Goldberg, I. H. (1983) *Biochemistry 22*, 4872.

Kim, S. H., Sussman, J. L., & Church, G. M. (1974) *Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions* (Sundaralingam, M., & Rao, S. T., Eds.) pp 571–575, University Park Press, Baltimore, MD.

Kolchinskii, A. M., Mirazabekov, A. D., Zasedatelev, A. S., Gurskii, G. V., Grokhovskii, S. L., Zhuze, A. L., & Gottikh, B. P. (1975) *Mol. Biol. (Kiev) 9*, 14.

Kopka, M. L., Yoon, C., Goodsell, D., Pjurra, P., & Dickerson, R. E. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1376–1380.

Lane, M. J., Dabrowiak, J. C., & Vournakis, J. N. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3260–3264.

Laverly, R., & Pullman, B. (1985) *J. Biomol. Struct. Dyn. 2*, 1021–1032.
Lee, J. S., & Waring, M. J. (1978) *Biochem. J. 173*, 129.
Le Pecq, J.-B., & Paoletti, C. (1967) *J. Mol. Biol. 27*, 87–106.
Lewis, M., Jeffrey, A., Ladner, R., Ptashne, M., Wang, J., & Pabo, C. O. (1982) *Cold Spring Harbor Symp. Quant. Biol. 47*, 435.
Lown, J. W., & Joshua, A. V. (1982) *J. Chem. Soc., Chem. Commun.*, 1298–1300
Lown, J. W., & Krowicki, K. (1985) *J. Org. Chem. 50*, 3774–3779.
Lown, J. W., Plenkiewicz, J., Ong, C. W., Joshua, A. V., McGovren, J. P., & Hanka, L. J. (1984) *Proceedings of the 9th International Congress of Pharmacology*, pp 265–269, Macmillan, London.
Lown, J. W., Sondhi, S. M., Ong, C.-W., Skorbogaty, A., Kishikawa, H., & Dabrowiak, J. C. (1986a) *Biochemistry 25*, 5111–5117.
Lown, J. W., Krowicki, K., Balzarini, J., & De Clercq, E. (1986b) *J. Med. Chem. 29*, 1210–1214.
Maniatis, T., Fritschi, E. F., & Sambrook, J. (1982) *Molecular Cloning—A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Marky, L. A., Blumenfeld, K. S., & Breslauer, K. J. (1983) *Nucleic Acids Res. 11*, 2857–2870.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol 65*, 499–560.
McGhee, J. D. (1976) *Biopolymers 15*, 1345.
Morgan, A. R., Lee, J. S., Pulleyblank, D. E., Murray, N. L., & Evans, D. H. (1979) *Nucleic Acids Res. 7*, 547–569.
Neidle, S., & Waring, M. J., Eds. (1983) *Molecular Aspects of Anti-cancer Drug Action. Topics in Molecular and Structural Biology*, Vol. 3, pp 35–55, 127–156, 157–181, Macmillan, London.
Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., & Matthew, B. W. (1982) *Nature (London) 298*, 718.
Patai, S. (1975) *Chemistry of Amidines and Imidates*, Wiley, London.
Patel, D. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6424.
Pinner, A., & Klein, F. (1877) *Chem. Ber. 10*, 1889.
Reinert, K. E., & Thorson, H. (1970) *Stud. Biophys. 24/25*, 319.
Ruddon, R. W. (1981) *Cancer Biology*, Oxford University Press, Oxford.
Sobell, H. M., & Jain, S. C. (1972) *J. Mol. Biol. 68*, 21.
Takeda, Y., Ohlendorf, D. H., Anderson, W. F., & Matthews, B. W. (1983) *Science (Washington, D.C.) 221*, 1020.
Van Dyke, M. W., Hertzberg, R. P., & Dervan, P. B. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5470–5474.
von Hippel, P. H. (1979) in *Biological Regulation and Development* (Goldberger, R. F., Ed.) p 279, Plenum, New York.
Ward, B., Skorobogaty, A., & Dabrowiak, J. C. (1986) *Biochemistry* (in press).
Wartell, R. M., Larson, J. E., & Wells, R. D. (1974) *J. Biol. Chem. 249*, 6179.
Westhof, E., Prange, T., Chevier, B., & Moras, D. (1985) *Biochimie 67*, 811–817.
Winter, R. B., & von Hippel, P. H. (1981) *Biochemistry 20*, 6948–6960.
Winter, R. B., Berg, O. G., & von Hippel, P. H. (1981) *Biochemistry 20*, 6961.
Zimmer, C. H. (1975) *Prog. Nucleic Acid Res. Mol. Biol. 15*, 285.
Zimmer, C., Luck, G., & Thrum, H. (1979a) *Stud. Biophys. 24/25*, 311.
Zimmer, C., Marck, C., Schneider, C., & Guschlbauer, W. (1979b) *Nucleic Acids Res. 6*, 2831.