GC Sequence Specific Recognition by an N-Formamido, C-Terminus-Modified and Imidazole-Containing Analogue of Netropsin

Moses Lee,* Christopher S. Preti, Emily Vinson, Michael D. Wyatt,† and John A. Hartley†

Department of Chemistry, Furman University, Greenville, South Carolina 29613, and Department of Oncology, University College London Medical School, London, WIP 8BT, U.K.

This communication describes, to our knowledge, the first example of a lexitropsin that recognizes solely GC sequences of DNA, and it will have immediate application in the development of new pharmaceutical agents,[1] artificial restriction enzymes,[2] as well as DNA probes.[3] There is a current interest in the design of small molecules that can bind to the minor groove of DNA, especially to GC-rich sequences. This interest stems from the observation that most minor groove binders are AT sequence specific[4] and that GC-rich sequences are commonly found in genomes of mammals, including humans. A functional role of such GC-rich sequences is suggested by their frequent occurrence in genes associated with proliferation, including a number of oncogenes.[5]

Netropsin 1 (see Figure 1) and distamycin, members of a class of naturally occurring antibiotic oligopeptides that bind to the minor groove of $(A/T)_4$ and $(A/T)_5$ sequences, respectively,[6,7] have been widely employed as models for the design of GC sequence-selective lexitropsins.[8] It was shown that replacement of the pyrrole moieties of netropsin with imidazoles[8-10] and removal of one of the positive charges, such as in analogue 2,[10] permitted GC sequence recognition. Due to steric interactions between the methylene groups of 2 with the guanine-2-NH_2 group in the minor groove of a GC base pair, the C-terminus of this ligand was forced to bind to an AT base pair, thus explaining its selectivity for the sequence 5′-(G/C)3(A/T).[10] It was also demonstrated that a truncated netropsin analogue, 3, selectively binds to the sequence 5′-ATTG-3′ when challenged with d(CGCAATTGCG)₂, indicating that the modified C-terminus could accept a GC base pair.[11] While the design of GC sequence-selective lexitropsins,[9b,c] such as 2,[10] was viewed as a significant achievement, Kopka and Larsen stated that the development of netropsin analogues which can recognize pure GC sequences still remains a formidable task.[8c]

In this communication we report the DNA binding studies of an imidazole-containing analogue, 4[12] (Figure 1). The N- and C-termini of this compound contain a formamido and a carboxamido moiety, respectively, replacing the guanidiniumylacetyl and propanamidinium groups of netropsin, and it represents a departure from our previous prototype compounds such as 2. The inclusion of a basic (dimethylamino)ethyl moiety (pK_a ~9.8) at the -terminal imidazole of 4 ensured that it would be protonated at a physiological pH of 7.4, thus making it monocationic. On the basis of recent reports
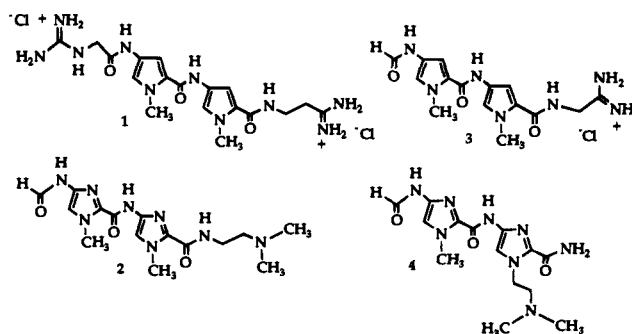

Figure 1. Structures of netropsin 1, the monocationic imidazole-containing lexitropsin 2, the truncated netropsin analogue 3, and the target compound 4.
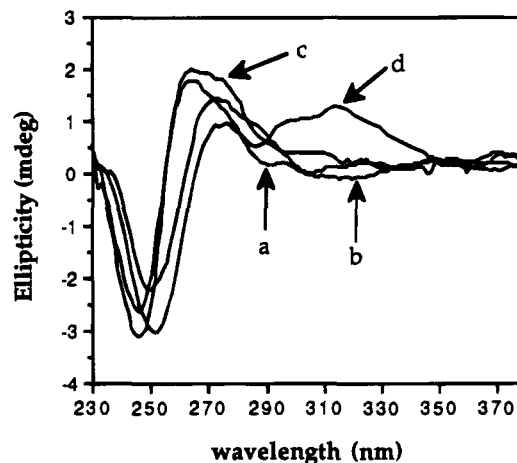

Figure 2. Representative overlaid CD spectra of the titration of compound 4 to poly(dA-dT) and poly(dG-dC). Spectra a and b are for free poly(dA-dT) and poly(dG-dC), respectively. Spectra c and d represent the titration of compound 4 to the respective DNAs at *r*′ values of 0.8.

on the DNA binding properties of *N*-pyrrole polyamine-substituted distamycin analogues, the positively charged dimethylammonium group of agent 4 should be positioned to electrostatically interact with the phosphate backbone thereby stabilizing the ligand:DNA complex.[13]

The apparent binding constants of analogue 4 to calf thymus DNA, T4 coliphage DNA, poly(dG-dG) and poly(dA-dT) determined using an ethidium displacement assay[14] were $(1.37 \pm 0.05) \times 10^5$, $(1.9 \pm 0.3) \times 10^4$, $(1.05 \pm 0.04) \times 10^5$, and $(6.5 \pm 0.5) \times 10^4$ M^{-1}, respectively. According to these data, analogue 4 is a weaker binder than netropsin 1 which has a binding constant to calf thymus DNA of 1.87×10^6 M^{-1} under identical conditions.[9a] The difference may result from several factors, including the smaller number of positively charged groups and van der Waals sites on 4 compared to those on 1. The results, however, show that compound 4 binds approximately twice as strongly to poly(dG-dC) than to poly(dA-dT). In comparison, netropsin binds approximately 1000-fold more strongly to poly(dA-dT) than to poly(dG-dC),[15] thus demonstrating the acceptance of GC base pairs by compound 4. The interaction of 4 to T4 coliphage DNA suggests that it interacts in the minor groove of DNA.[16a,b]

The binding of compound 4 to poly(dA-dT) and poly(dG-dC) was also studied by CD. The experiments were identically performed with regard to temperature and concentrations of drug and DNA at defined *r*′ values, the ratio of the moles of drug to DNA base pairs.

---

† University College London Medical School.

a b c d e f

— 550
— 540
— 530
— 520
— 510
— 500
— 490
— 480
— 470
— 460
— 450
— 440

431   [////]   [////] 451   ●●●●   471   oooo [////]

G C T G G C G C C T A T A T C G C C G A C A T C A C C G A T G G G G A A G A T C G G G C T C G C C A
C G A C C G C G G A T A T A G C G G C T G T A G T G G C T A C C C C T T C T A G C C C G A G C G G T

481 ooo   oooo 501 [////]   ●●●●●oooooo●●●●●   521   ●●

C T T C G G G C T C A T G A G C G G C T T G T T T C G G C G T G G G T A T G G T G G C A G G C C C G T
G A A G C C C G A G T A C T C G C G A A C A A A G [C C G C] A C C C A T A C C A C C G T [C C G G] G C A

531 ●●   oooo   551   571

G G C C G G G G G A C T G T T G G G C G C C A T C T C C T T G C A T G C A C C A T T C C T T G C G G
[C C G G] C C C C C T G A C A A C C C G C G G T A G A G G A A C G T A C G T G G T A A G G A A C G C C

Figure 3. (a, Left) A photograph of the autoradiogram of the MPE-Fe(II) footprinting of compounds **2** and **4** on the *Bam*H1-*Sal*1 fragment of pBR322 plasmid DNA singly 5′-^{32}P-end labeled at the *Bam*H1 site. Lane a, control unmodified DNA; lane b, formic acid (G + A) marker lane; lane c, compound **2** (250 µM); lane d, control MPE cut; lanes e and f, compound **4** (250 and 500 µM, respectively). (b, Right) Densitometric analysis of the MPE footprinting autoradiograms on the sequenced portion of the *Bam*H1-*Sal*1 fragment of pBR322 plasmid DNA. Boxes and circles above the sequence represent the *Bam*H1 end-labeled fragment, and boxes below represent the *Sal*1 end-labeled fragment. Shaded boxes represent weak footprints of compound **4**. Open and filled circles represent the weak and strong footprints of compound **2**, respectively.

Titration of compound **4** to poly(dA-dT) did not produce any DNA-induced ligand bands even at an *r*′ of 1.0; however, addition of this compound to poly(dG-dC) gave rise to a positive Cotton effect at 318 nm of 0.5 mdeg at an *r*′ of 0.3. As shown in Figure 2 (trace d), when the *r*′ value was raised to 0.8 the ellipticity was 1.5 mdeg. The induced band is presumably due to the UV π to π* absorption of the drug in the drug:DNA complex.[4a] Since compound **4** itself did not produce a CD spectrum, the appearance of a DNA-induced ligand band was taken as evidence for its interaction with DNA. In accordance with previous reports,[4a,17] and assuming that compound **4** binds to the polynucleotides in a similar orientation, the above data suggest that the compound exhibits selectivity for poly(dG-dC) over poly(dA-dT). In the titration studies the bands at 240 and 265 nm are consistent with those of B-DNA, thus suggesting that binding of ligand **4** only causes minimal distortions in the DNA conformation.

In order to ascertain the DNA sequence selectivity of compound **4**, MPE-Fe(II) footprinting studies were performed, using a 5′-^{32}P-singly end-labeled *Bam*H1-*Sal*1 fragment of pBR322 plasmid DNA. A representa-

tive gel is shown in Figure 3a. Analysis of the resolvable portion of the autoradiogram by densitometry (see Figure 3b) revealed that the reaction of **4** gave weak but distinct footprints covering four base pairs at 5′-C(436)GCC(439), 5′-G(446)CCG(449), 5′-C(476)GCC-(479), and 5′-C(505)GGC(508) sequences on the fragment labeled at the *Bam*H1 site. When the *Sal*1-labeled DNA fragment was analyzed, the following sites were inhibited from cleavage by MPE: 5′-C(509)GCC(506), 5′-C(529)GGG(526), 5′-G(534)GCC(531), 5′-G(552)GCG-(549), 5′-C(580)CGC(577). All of the footprint sites mapped occur within pure GC sequences with total avoidance of the AT sites present in this DNA fragment, thus, demonstrating the GC sequence specificity of compound **4**.

As with other lexitropsins,[9,10] binding of agent **4** to four contiguous base pairs is in accordance with Dervan's N + 1 rule for the binding density of netropsin analogues, in which N is the number of amido groups. It is, thus, unlikely that any part of compound **4** would be intercalated into the DNA because the binding density would otherwise be smaller than four.[18] The conclusion is also supported by the similar DNA binding

constants of compound **4** to those of **2**, a known minor groove binder.[10] The binding of agent **4** to the minor groove implies that the amido protons should be able to form bifurcated hydrogen bonds with guanine-N3 and cytosine-O2 in the floor of the minor groove to stabilize the ligand:DNA complex. In addition, the strict recognition of pure GC sequences also suggests that the imidazole-N3 atoms on the concave face of compound **4** could provide both the space for accomodating and hydrogen bonding with the exocyclic guanine-2-NH$_2$ group. As a control in the footprinting experiment, compound **2** (Figure 3a, lane c) produced strong footprints at other sites on the same fragment of DNA that contain the expected 5'-(G/C)$_3$(A/T)-3' sequence.[10] Further studies on interactions of compound **4** with oligodeoxyribonucleotides by ^{1}H-NMR and X-ray studies are in progress, and the results will be reported in due course.

Supplementary Material Available: The synthetic scheme for the preparation of compound **4**, its ^{1}H-NMR spectrum as well as the densitometric scans of the MPE footprinting experiments are given (4 pages). Ordering information is given on any current masthead page.

References

(1) (a) *Molecular Aspects of Anti-Cancer Drug Action*; Neidle, S., Waring, M., Eds.; CRC Press: Boca Raton, 1993; Vol. 1. (b) *Molecular Basis of Specificity in Nucleic Acid-Drug Interactions*; Pullman, B., Jortner, J., Eds.; Kluwer Academic Publishers: Boston, 1990.

(2) Dervan, P. B. Design of Sequence-Specific DNA Binding Molecules. *Science* **1986**, *232*, 464–471.

(3) Bains, W. Disease, DNA and Diagnosis. *New Scientist* **1989**, May 6, 48–51.

(4) (a) Zimmer, C.; Wahnert, U. Nonintercalating DNA Binding Ligands: Specificity of Interaction and Their Use as Tools in Biophysical, Biochemical and Biological Investigations of Genetic Material. *Progr. Biophys. Mol. Biol.* **1986**, *47*, 31–112. (b) *Advances in DNA Sequence Specific Agents*; Hurley, L., Ed.; JAI Press: Greenwich, CT, 1992, Vol. 1. (c) Thurston, D. E.; Thompson, A. S. The Molecular Recognition of DNA. *Chem. Br.* **1990**, 767–772.

(5) Barbacid, M. *ras* Genes. *Annu. Rev. Biochem.* **1987**, *56*, 779–872.

(6) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. The Molecular Origin of DNA-Drug Specificity of Netropsin and Distamycin. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376–1380.

(7) Hahn, F. E. Distamycin A and Netropsin. In *Antibiotics III. Mechanism of Action of Antimicrobial and Antitumor Agents*; Corcoran, J. C., Hahn, F. E., Eds.; Springer-Verlay: New York, 1975; pp 79–100.

(8) (a) Krowicki, K.; Lee, M.; Hartley, J. A.; Ward, B.; Kissinger, K.; Skorobogaty, A.; Dabrowiak, J. C.; Lown, J. W. Molecular Recognition between Oligopeptides and Nucleic Acids – Rational Design of Sequence Specific DNA Binding Agents. In *Structure and Expression*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: Albany, 1988; Vol. 2, pp 251–271. (b) Lown, J. W. Lexitropsins: Rational Design of DNA Sequence Reading Agents as Novel Anti-Cancer Agents and Potential Cellular Probes. *Anti-Cancer Drug Des.* **1988**, *3*, 25–40. (c) Kopka, M. L.; Larsen, T. A. Netropsin and Lexitropsins. The Search for Sequence-Specific Minor-Groove-Binding Ligands. In *Nucleic Acid Targeted Drug Design*; Propst, C. L., Perun, T. J., Eds.; Marcel Dekker, Inc.: New York, 1992; pp 303–374.

(9) (a) Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C. Molecular Recognition between Oligopeptides and Nucleic Acids: Novel Imidazole-Containing Oligopeptides Related to Netropsin that Exhibited Altered DNA Sequence Specificity. *Biochemistry* **1986**, *25*, 7408–7416. (b) Kissinger, K.; Krowicki, K.; Dabrowiak, J. C.; Lown, J. W. Molecular Recognition between Oligopeptides and Nucleic Acids: Monocationic Imidazole Lexitropsins that Display Enhanced GC Sequence Dependent Reading. *Biochemistry* **1987**, *26*, 5590–5598. (c) Lee, M.; Hartley, J. A.; Pon, R. T.; Krowicki, K.; Lown, J. W. Sequence Specific Molecular Recognition by a Monocationic Lexitropsin and the Decadeoxyribonucleotide d(CATGGCCATG)$_2$. Structural and Dynamics Properties Deduced from High Field ^{1}H-NMR Studies. *Nucl. Acids Res.* **1988**, *16*, 665–684.

(10) Lee, M.; Rhodes, A. L.; Wyatt, M. D.; Forrow, S.; Hartley, J. A. GC Base Sequence Recognition by Oligoimidazolecarboxamide and C-Terminus Modified Analogues of Distamycin Deduced from CD, ^{1}H-NMR and MPE Footprinting Studies. *Biochemistry* **1993**, *32*, 4237–4245.

(11) Lee, M.; Krowicki, K.; Hartley, J. A.; Pon, R. T.; Lown, J. W. Molecular Recognition between Oligopeptides and Nucleic Acids: Influence of van der Waals Contacts in determining the 3'-Terminus of DNA Read by Monocationic Lexitropsins. *J. Am. Chem. Soc.* **1988**, *110*, 3641–3649.

(12) The synthetic scheme for the preparation of compound **4** is given in the supplementary material section: mp 170 °C dec; TLC (10% MeOH/CHCl$_3$) R_f 0.77; ^{1}H-NMR (CDCl$_3$) δ 9.18 (br s 1H), 8.36 (s, 1H), 7.60 (br s, 1H), 7.55 (s, 1H), 7.46 (s, 1H), 7.05 (br s, 1H), 5.29 (br s, 1H), 4.57 (t, 7.2, 2H), 4.08 (s, 3H), 2.73 (t, 7.8, 2H), 2.30 (s, 6H); IR (CHCl$_3$ cast) ν 3412, 2921, 1665, 1534, 1458, 1376 cm^{-1}; UV (water) λ_{max} 205, 302 nm (ϵ = 104 950, 35 019 M^{-1} cm^{-1}); MS (FAB, NBA/TFA) m/z (intensity) 349 (34, M + H); FAB-HRMS m/z 349.1804 (C$_{14}$H$_{20}$N$_8$O$_3$ + H), calcd 349.1787.

(13) (a) He, G.-X.; Browne, K. A.; Groppe, J. C.; Blasko, A.; Mei, H.-Y.; Bruice, T. C. *Microgonotropens* and Their Interactions with DNA. 1.[1] Synthesis of the Tripyrrole Peptides Dien-Microgonotropen-a, -b, and -c and Characterization of Their Interactions with dsDNA. *J. Am. Chem. Soc.* **1993**, *115*, 7061–7071. (b) Blasko, A.; Browne, K. A.; He, G.-X.; Bruice, T. C. Microgonotropens and Their Interactions with DNA. 3.[1] Structural Analysis of the 1:1 Complex of d(CGCAATTGCG)$_2$ and Dien-Microgonotropen-c by 2D-NMR Spectroscopy and Restrained Molecular Modeling. *J. Am. Chem. Soc.* **1993**, *115*, 7080–7092.

(14) Morgan, A. R.; Lee, J. S.; Pulleyblank, D. E.; Murray, N. L.; Evans, D. H. Ethidium Fluorescence Assays. Part 1. Physiochemical Studies. *Nucl. Acids Res.* **1979**, *7*, 547–569.

(15) Rao, K. E.; Shea, R. G.; Yadagiri, B.; Lown, J. W. Molecular Recognition between Oligopeptides and Nucleic Acids: DNA Sequence Specificity and Binding Properties of Thiazole-Lexitropsins Incorporating the Concepts of Base Site Acceptance and Avoidance. *Anti-Cancer Drug Des.* **1990**, *5*, 3–20.

(16) (a) Lown, J. W. New Approaches to the Study of the Mechanism of Action of Antitumor Antibiotics *Acc. Chem. Res.* **1982**, *15*, 381–387. (b) The ellipticity of the DNA induced ligand band of the poly(dG-dC):4 complex (r' = 0.8) was reduced by the addition of NaCl (0.76–13.9 μM final concentrations). The dependence of binding on ionic strength suggests that the interaction of compound **4** with DNA is in part controlled by electrostatic forces, which is in agreement with groove binding (see: Lee, M.; Shea, R. G.; Hartley, J. A.; Kissinger, K.; Pon, R. T.; Vesnaver, G.; Breslauer, K. J.; Dabrowiak, J. C.; Lown, J. W. Molecular Recognition between Oligopeptides and Nucleic Acids: Sequence Specific Binding of the Naturally Occurring Antibiotics (4S)-(+)-Anthelvencin A and Its (4R)-(-) Enantiomer to Deoxyribonucleic Acids Deduced from ^{1}H-NMR, Footprinting, and Thermodynamic Data. *J. Am. Chem. Soc.* **1989**, *111*, 345–354).

(17) Burckhardt, G.; Luck, G.; Strol, K.; Zimmer, C.; Lown, J. W. Binding to DNA of Selected Lexitropsins and Effects on Procaryotic Topoisomerase Activity. *Biochim. Biophys. Acta* **1993**, *1173*, 266–272.

(18) Thus far (see refs 8–10) no imidazole-containing analogues of netropsin has been shown to intercalate DNA. Of all the netropsin analogues studied only one thiazole-containing compound was demonstrated to bind in a combination of intercalation and minor groove binding modes (Kumar, S.; Jaseja, M.; Zimmerman, J.; Yadagiri, B.; Pon, R. T.; Saspe, A.-M.; Lown, J. W. *J. Biomol. Struct. Dyn.* **1990**, *8*, 99–121).