Pergamon

Bioorganic & Medicinal Chemistry Letters, Vol. 4, No. 6, pp. 801-806, 1994

0960-894X/94 $6.00+0.00

0960-894X(94)E0054-I

SYNTHESIS AND DNA BINDING PROPERTIES OF A SERIES OF N TO C LINKED AND IMIDAZOLE CONTAINING ANALOGUES OF DISTAMYCIN

Michael D. Wyatt,[a] Bonnie J. Garbiras,[b] Moses Lee,[b]* Stephen M. Forrow,[a] John A. Hartley[a]

[a]Department of Oncology, UCL Medical School, London W1P 8BT U. K.,

[b]Department of Chemistry, Furman University, Greenville S.C. 29613

Abstract- A series of imidazole containing analogues linked in a head to tail fashion (N to C) and containing methylene linkers were synthesized and examined for their DNA binding properties. The triimidazole β-alanine linked compound **10a** was shown to bind efficiently to GC rich sites on DNA covering six base pairs, which is in contrast to the tetraimidazole carboxamide compound **3**, previously shown to be a poor DNA binder.

There is widespread interest in the development of reagents which recognize and bind to DNA in a sequence specific manner. One ultimate goal in such an approach would be the ability to construct ligands designed to bind specifically to any chosen DNA sequence.[1] The naturally occurring oligopeptides netropsin **1** and distamycin **2** (see Figure 1), which bind to the minor groove of DNA and recognize (A/T)4 and (A/T)5 respectively, have proven particularly useful in the gaining of insights into ligand:DNA interactions.[2] It has been shown that replacement of the pyrrole heterocycle with an imidazole creates compounds which allow the recognition of GC base pairs.[3] It also has been shown that attempting to increase the binding site size by increasing the number of pyrrole carboxamide units results in compounds with a decreased binding affinity to DNA compared to **1** and **2**.[4] This is most likely a combination of two problems, those of phasing[5] and of curvature.[6]

It has recently been demonstrated clearly with respect to imidazole analogues that there is a marked decrease in the binding affinity of a tetraimidazole compound **3**, with a theoretical binding site size of six base pairs,[4] compared to its triimidazole counterpart **4** (binding to five base pairs). This was rationalized to be due to the increased radius of curvature of the tetraimidazole compound, thereby preventing maximum hydrogen bonding and van der Waals forces between the ligand and DNA.[7] The concept of utilizing different linkers to overcome the problems encountered with carboxamide linked oligopyrrole analogues of **1** and **2** has been explored. In work describing the linking of netropsin units in three different manners, N to N (head to head)[8,9], C to C (tail to tail)[8] and N to C (head to tail)[8,10], the compounds that were synthesized were shown to span between ten and eleven base pairs in a bidentate manner.

In the present study a series of imidazole compounds linked in the head to tail sense were synthesized and studied by an ethidium assay, CD, MPE footprinting and molecular modelling studies. The aliphatic linker consists of either two (β-alanine) or three (γ-aminobutyrate) methylene groups which tether different numbers of imidazole units and were designed to explore the optimum binding of small molecules to GC rich sequences of increasing size. This interest stems from the observation that regions of high GC content are commonly found in genomes of mammals, including humans, and that a functional role of GC rich sequences is suggested by their frequent occurrence in genes associated with proliferation including a number of oncogenes.[11,12]

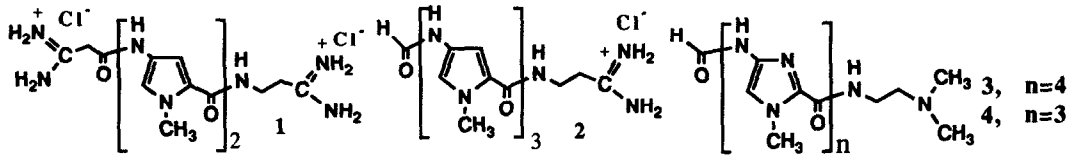
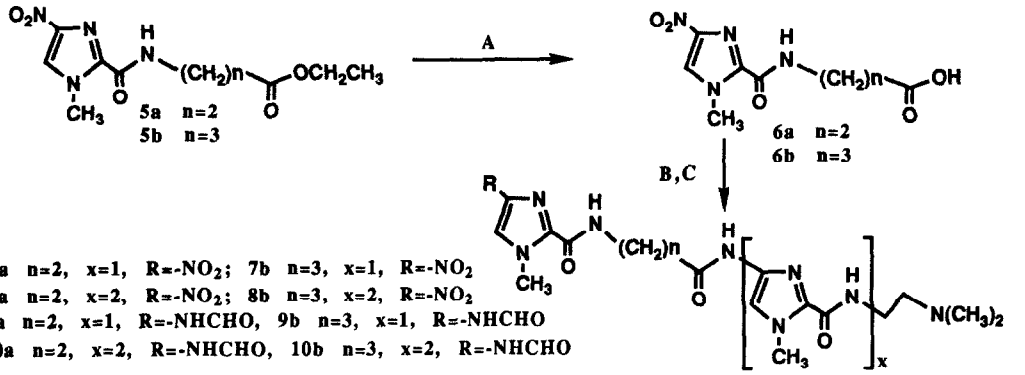


Figure 1. Netropsin **1**, distamycin **2**, and the corresponding GC recognizing tetraimidazole compound **3**, and triimidazole compound **4**.

Synthesis. Reaction of 1-methyl-4-nitroimidazole-2-carbonyl chloride[13] with β-alanine ethyl ester hydrochloride gave the ethyl ester **5a** in 96% yield (see Scheme 1). The ester was then hydrolyzed to produce the carboxylic acid **6a** which was coupled to N-(N',N'-dimethylaminoethyl)-4-amino-1-methylimidazole-2-carboxamide[7] in the presence of dicyclohexylcarbodiimide (DCC) and dimethylaminopyridine (DMAP) to give the linked analogue **7a** as an off white powder in 38% yield. The coupling of the acid **6a** with N-(N',N'-dimethylaminoethyl)-1-methyl-4-[4-amino-1-methylimidazole-2-carboxamido]imidazole-2-carboxamide[7] was performed under identical conditions and **8a** was isolated in 35% yield. The syntheses of the γ-aminobutyrate compounds were performed in an identical manner, except γ-aminobutyrate hydrochloride was used (see Scheme 1). All four nitroimidazole compounds, **7a**, **7b**, **8a**, and **8b**, were then catalytically reduced to the corresponding amines and reacted with acetic formic anhydride to produce the formamido compounds **9a** (39%, 126°C (dec)), **9b** (26%, 128°C(dec)), **10a** (33%, 134-139°C), and **10b** (71%, 169-172°C), respectively.[14] The structures and purity of all these compounds were ascertained by NMR and high resolution mass spectral analyses.



7a n=2, x=1, R=-NO_2; **7b** n=3, x=1, R=-NO_2
8a n=2, x=2, R=-NO_2; **8b** n=3, x=2, R=-NO_2
9a n=2, x=1, R=-NHCHO, **9b** n=3, x=1, R=-NHCHO
10a n=2, x=2, R=-NHCHO, **10b** n=3, x=2, R=-NHCHO

Scheme 1. A 1) NaOH, 2)HCl; B N-(N',N'-dimethylaminoethyl)-4-amino-1-methylimidazole-2-carboxamide or N-(N',N'-dimethylaminoethyl)-1-methyl-4-[4-amino-1-methylimidazole-2-carboxamido]imidazole-2-carboxamide, DCC, DMAP, CH_2Cl_2; C 1)H_2, Pd-C, methanol, 2) acetic formic anhydride, CH_2Cl_2.

DNA Binding Properties. The apparent binding constants of the compounds to four representative DNAs of differing GC content were determined using an ethidium bromide assay[15] and are reported in Table 1. Binding to T4 coliphage DNA suggests minor groove interaction because the major groove of the DNA is occluded by α-glycosylation of the cytidine residues. Several trends are evident from the data. Firstly, the β-alanine linked

analogues (**9a** and **10a**) have higher binding affinities than the γ-aminobutyrate analogues, implying that the compounds with the $(CH_2)_2$ linker have more favorable interactions with the DNAs studied than those with three methylenes. Secondly, the triimidazole analogues (**10a** and **10b**) bind more strongly compared to the corresponding diimidazole analogues (**9a** and **9b**), most likely due to the extra amido group and larger van der Waals interactions in the case of the triimidazole compounds. Also, the β-alanine linked analogues have a markedly improved acceptance of GC base pairs, most evident with the triimidazole compound **10a**, which also has a higher binding constant to calf thymus and T4 DNA than the AT specific natural product distamycin **2**.

Table 1. Association Constants (K_{app}, x 10^5 M^{-1}) of the Compounds with Polynucleotides.

Compound	Calf Thymus	T4	Poly(dA-dT)	Poly(dG-dC)
2[7]	7.7	6.5	348	2.0
9a	1.5±0.1	1.4±0.1	0.12±0.04	5.8±0.2
9b	0.10±0.05	0.33±0.03	0.10±0.04	0.21±0.06
10a	16.0±0.2	19.0±0.5	2.5±0.5	52.0±0.5
10b	0.48±0.1	1.4±0.2	0.50±0.1	0.92±0.2

Circular Dichroism. The results from CD titration studies show that all of the compounds bind to the DNAs studied as indicated by the appearance of ligand-induced bands between ~295 and 305 nm.[16] Because the compounds do not exhibit any CD spectra by themselves this is taken as clear evidence of their interactions with the DNAs.[17] Titration of the β-alanine linked analogue **9a** with poly(dA-dT) produced a positive DNA induced ligand Cotton effect at 295 nm with a strength of 1.5 mdeg at an r' value of 0.25. The r' values correspond to the moles of ligand to the moles of DNA base pairs. The effect of **9a** on poly(dG-dC) was greater as the positive DNA induced ligand band seen at 295 nm showed an ellipticity of 4.5 mdeg at the same r' value, over twice the effect produced by binding to poly(dA-dT). For **10a**, the larger effect was also seen on poly(dG-dC) (see Figure 2a), with an ellipticity value of 5.3 mdeg at an r' value of 0.25, while the value for poly(dA-dT) was 3.0 mdeg at the same r' value (see Figure 2b). The data, which are in agreement with the Kapp values, demonstrate the stronger binding affinity of the β-alanine linked analogues **9a** and **10a** to poly(dG-dC) over poly(dA-dT).

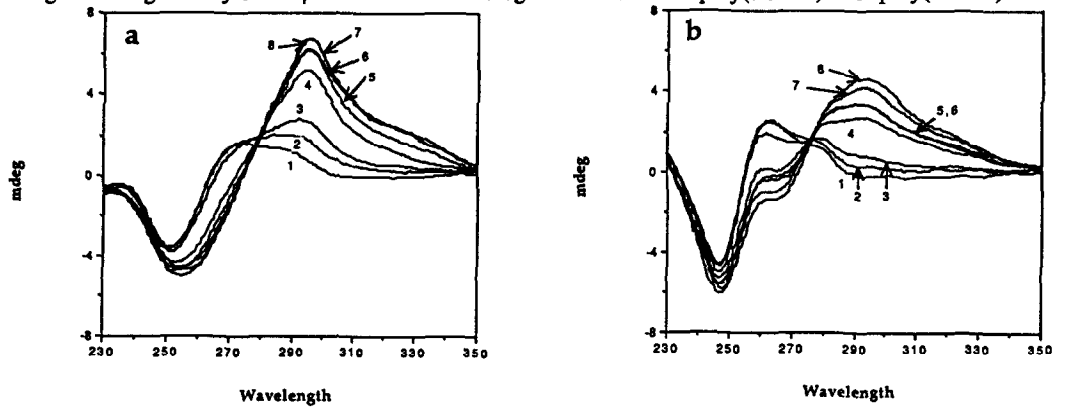


Figure 2. Titration of **10a** to (a) poly(dG-dC) and (b) poly(dA-dT). The spectra correspond to r' values of 0, 0.05, 0.12, 0.25, 0.35, 0.5, 0.6, 0.7 for part (a), and 0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35 for part (b).

Similar titration studies were performed with the γ-aminobutyrate linked compounds. The appearance of a ligand induced band at 295 nm was evident upon titration of analogues **9b** and **10b** with the DNAs, though the interactions proved weaker than with **9a** and **10a** (data not shown). In all these experiments, the retention of the native negative and positive bands for B-DNA suggest that the binding of the molecules to the DNAs are not significantly altering the DNA structure from B-form.

MPE Footprinting. The four compounds **9a**, **10a**, **9b**, and **10b** were examined on the 274 base pair GC rich Sal I/ Bam HI fragment of pBR322 DNA (see Figure 3). The γ-aminobutyrate linked compounds gave poor footprints at 100-500 μM, while the β-alanine linked compounds provided several clear protection sites from cleavage under identical conditions. Compounds **9a** and **10a** possessed two common sites of protection, (585-590) 5'-(G)GTGCT-3' and (580-575) 5'-(G)CCGCA-3', with the extra base pair recognized by **10a**. In addition, there were differing sites of recognition for the compounds. A five base pair site recognized by **9a**, but only weakly by **10a** consisted of the recognition site within the sequence (545-539) 5'-AACAGTC-3'. Two additional sites recognized by **10a** were (525-530) 5'-GCCCGT-3' and (569-564) 5'-GGTGCA-3', of which the latter proved the strongest binding site for **10a** in the DNA fragment studied. The sequenced portion of this DNA fragment contains sites with six (524-529, 546-551) or more (531-539, 577-586) contiguous GC base pairs, but **9a** and **10a** did not bind to sites that contained only GC base pairs in the recognition sequences. The compounds therefore clearly recognize GC rich sequences with the avoidance of AT rich sequences and show a tolerance for the isolated AT base pair within binding sites, possibly due to the linker, as well as the 3'-AT base pair, shown previously to be due to the methylene groups on the C-terminus of the molecules.[18]

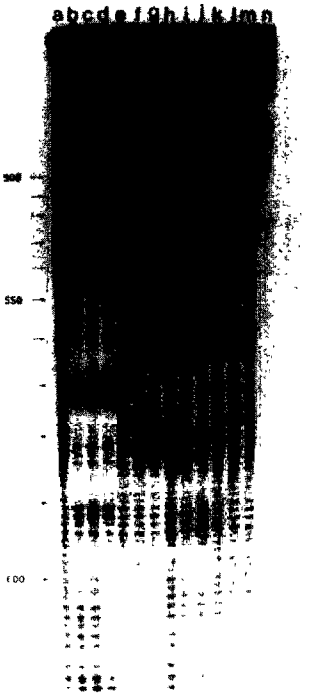


Figure 3. Autoradiogram of the MPE footprinting gel on the Sal I-labeled Sal I/Bam HI fragment. Lane a, MPE alone; lanes b-d, **10a**, 100, 250, 500 μM, respectively; lanes e-g **10b**, 100, 250, 500 μM, respectively; lanes h-j, **9a**, 100, 250, 500 μM, respectively; lanes k-m, **9b**, 100, 250, 500 μM, respectively; lane n, formic acid (G+A) lane.

For these β-alanine compounds, the approximated binding site size corresponded to 5 and 6 base pairs for **9a** and **10a**, respectively, suggesting binding of both arms of the molecule (bidentate binding). The triimidazole compound **10a** is therefore binding to the theoretical site size of compound **3** as predicted by the n+1 rule[4], but at a much greater affinity than **3** which does not footprint to the same DNA fragment under identical conditions.[7]

Molecular Modelling. Conformational geometry analyses on compounds **10a** and **10b** by MM2 were performed on a CAChe Tektronix system in an attempt to rationalize the differences observed in their binding affinities. The results showed that in both compounds the N3 atoms of the imidazole groups are pointing toward the DNA. Both conformers were docked onto the underlined sequence 5'-CCCCCAT with the C-terminal dimethylamino group located on the A residue. The positioning of the C-terminus over the AT site was based on previous studies.[7] The **10a**:heptamer complex was energy minimized with the amido NH groups of **10a** constrained to forming bifurcated hydrogen bonds to purine-N3 and pyrimidine-O2 sites. The resulting complex was minimized to 250 kcal/mol using MM2 on a conjugated gradient protocol. The results depicted in Figure 4 show that the ligand is snugly bound to the minor groove of the DNA with the proper hydrogen bonds in place. The repeat distance 'd' between the amides of the linker connecting the imidazoles was determined to be 4.5 Å. The radius of curvature of the ligand in the complex, determined using the concave imidazole-N atoms, is 19.3 Å which is compatible to the 19 Å demonstrated for B-DNA.[7] Furthermore, the linker resided over the third GC base pair 5' to the terminal AT base pair, and this offers an explanation for the appearance of an AT base pair at that position on several of the footprinting sites. Similar docking studies were done on **10b**, the results of which revealed that, when all the amido NH groups were constrained to forming hydrogen bonds with the purine-N3 and pyrimidine-O2 sites, the bond angles of the linker were significantly distorted. This implies that the γ-aminobutyrate linker is unable to adjust to an optimal repeat distance for binding properly in the minor groove of DNA and offers an explanation of its poor binding affinity. The studies indicate that linking imidazole units head to tail can produce compounds which bind to longer GC rich sequences. The concept of linkers offers promise in overcoming the phasing and curvature problems associated with simple carboxamide linked oligoimidazole compounds.

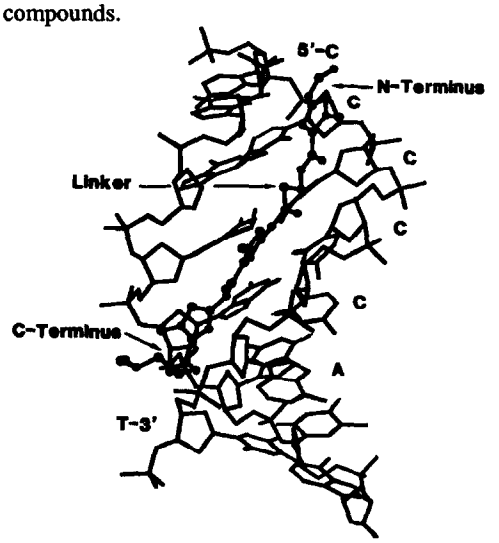


Figure 4. MM2 energy minimized 1:1 complex of **10a** on the underlined sequence of 5'-CCCCCAT.

Acknowledgements: This publication was supported in part by grants from the National Cancer Institute (no. 1 R15 CA56906-01), Research Corporation (no. C-2897) and the National Science Foundation-REU (no. 9000726). M.D.W. thanks the Consolidated Postgraduate Fund of University College London for a studentship.

References and Notes:

1. Neilson, P. E. *Bioconjugate Chem.* **1991**, *2*, 1.
2. Lown, J. W. *Anti-Cancer Drug Des.* **1988**, *3*, 25.
3. Kissinger, K.; Krowicki, K.; Dabrowiak, J. C.; Lown, J. W. *Biochem.* **1987**, *26*, 5590.
4. Youngquist, R. S.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2565.
5. Goodsell, D.; Dickerson, R. E. *J. Med. Chem.* **1986**, *29*, 727.
6. Cory, M.; Tidwell, R. R.; Fairley, T. A. *Ibid* **1992**, *35*, 431.
7. Lee, M.; Rhodes, A. L.; Wyatt, M. D.; Forrow, S.; Hartley, J. A. *Biochem.* **1993**, *32*, 4237.
8. Gursky, G. V.; Zasedatelev, A. S.; Zhuze, A. L.; Khorlin, A. A.; Grokhovsky, S. L.; Streltsov, S. A.; Surovaya, A. N.; Nikitin, S. M.; Krylov, A. S.; Retchinsky, V. O.; Mikhailov, M. V.; Beaubealashivili, R. S.; Gottikh, B. P. *Cold Spring Harbor Symp. Quant. Biol.* **1982**, *47*, 367.
9. (a) Lown, J. W.; Krowicki, K.; Balzarini, J.; Newman, R. A.; De Clercq, E. *J. Med. Chem.* **1989**, *32*, 2368. (b) Kissinger, K.; Dabrowiak, J. C.; Lown, J. W. *Chem. Res. Toxicol.* **1990**, *3*, 162. (c) Rao, K. E.; Zimmerman, J.; Lown, J. W. *J. Org. Chem.* **1991**, *56*, 786.
10. (a) Dasgupta, D.; Parrak, P. Sasisekharan, V. *Biochem.* **1987**, *26*, 6381. (b) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 7564.
11 Hartley, J. A.; Lown, J. W.; Mattes, W. B.; Kohn, K. W. *Acta. Oncol.* **1988**, *27*, 503.
12. Mattes, W. B.; Hartley, J. A.; Kohn, K. W.; Matheson, D. W. *Carcinogenesis* **1988**, *9*, 2065.
13. Krowicki, K.; Lown, J. W. *J. Org. Chem.* **1987**, *52*, 3493.
14. Data for **9a**. ^{1}H NMR ($CDCl_3$): δ 2.29 (s, 6H), 2.53 (t, 6.3, 2H), 2.64 (t, 6.0, 2H), 3.45 (q, 6.3, 2H), 3.73 (q, 6.0, 2H), 3.97 (s, 3H), 3.98 (s, 3H), 7.32 (s, 1H), 7.33, (s, 1H), 7.50 (br s, 1H), 7.58 (br s, 1H), 8.20 (br s, 1H), 8.22 (br s, 1H), 8.31 (s, 1H); IR(Nujol): ν 3390, 2953, 2840, 1654, 1540, 1016 cm^{-1}; UV (H_2O): λ_{max} 282 nm (ε 24320 cm^{-1} M^{-1}); HRMS (FAB, NBA) calcd. for $C_{18}H_{28}N_9O_4$ 434.2264. Found 434.2248 (M+H). For **9b**. ^{1}H NMR ($CDCl_3$): δ 2.06 (quintet, 6.3, 2H), 2.35 (s, 6H), 2.44 (t, 6.2, 2H), 2.62 (t, 6.2, 2H), 3.27 (m, 2H), 3.49 (q, 6.2, 2H), 3.99 (s, 3H), 4.01 (s, 3H), 7.33 (s, 1H), 7.39 (s, 1H), 7.61 (br s, 1H), 8.25 (br s, 1H), 8.34 (s, 1H), 8.45 (br s, 1H), 8.85 (br s, 1H); IR (Nujol): ν 3400, 2900, 1654, 1540, 668 cm^{-1}; UV (H_2O): λ_{max} 282 nm (ε 18468 cm^{-1} M^{-1}); HRMS (FAB, NBA/TFA) calcd. for $C_{19}H_{30}N_9O_4$: calcd. 448.2421. Found 448.2435 (M+H). For **10a**. ^{1}H NMR ($CDCl_3$): δ 2.19 (s, 6H), 2.47 (t, 6.0, 2H), 2.70 (t, 6.0, 2H), 3.45 (q, 6.0, 2H), 3.69 (q, 5.6, 2H), 3.89 (s, 3H), 3.90 (s, 6H), 7.28 (s, 3H), 7.81 (br s, 1H), 7.93 (br s, 1H), 8.13 (br s, 1H), 8.24 (s, 1H), 9.28 (br s, 1H), 9.62 (br s, 1H); IR(Nujol): ν 3396, 2932, 2846, 1649, 1536, 1211 cm^{-1}; UV (ethanol): λ_{max} 298 nm (ε 20367 cm^{-1} M^{-1}); HRMS (FAB-NBA/TFA) calcd. for $C_{23}H_{33}N_{12}O_5$ 557.2697. Found 557.2692 (M+H). For **10b**. ^{1}H NMR ($CDCl_3$): δ 2.02 (quintet, 7.2, 2H), 2.28 (s, 6H), 2.45 (t, 7.8, 2H), 2.50 (t, 6.3, 2H), 3.47 (q, 6.3, 2H), 3.49 (q, 6.3, 2H), 4.02 (s, 1H), 4.03 (s, 1H), 4.04 (s, 1H), 7.39 (s, 1H), 7.40 (s, 1H), 7.41 (s, 1H), 7.55 (s, 1H), 7.79 (s, 1H), 8.23 (s, 1H), 8.34 (s, 1H), 8.55 (s, 1H), 9.21 (s, 1H); IR (Nujol): ν 3400, 2800, 1649, 1542, 666 cm^{-1}; UV (ethanol): λ_{max} 294 nm (ε 15765 cm^{-1} M^{-1}); HRMS (FAB-NBA/TFA) calcd. for $C_{24}H_{35}N_{12}O_5$ 571.2853. Found 571.2822 (M+H).
15. (a) Morgan, A. R.; Lee, J. S.; Pulleybank, D. F.; Murray, N. L.; Evans, D. H. *Nucleic Acids Res.* **1979**, *7*, 547. (b) Debart, F.; Periguad, C.; Gosselin, D.; Mrani, D.; Rayner, B.; Le Ber, P.; Auclair, C.; Balzarini, J.; De Clercq, E.; Paoletti, C.; Imbach, J.-L. *J. Med. Chem.* **1989**, *32*, 1074.
16. The CD titration studies on poly(dA-dT) and poly(dG-dC) were performed using similar number of moles of DNA base pairs and range of r' values; therefore, the results are directly comparable.
17. (a) Burckhardt, G.; Luck, G.; Zimmer, C.; Storl, J.; Krowicki, K.; Lown, J. W. *Biochim. Biophys. Acta* **1989**, *1009*, 11. (b) Zimmer, C.; Wahnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31.
18. Lee, M.; Hartley, J. A.; Pon, R. T.; Krowicki, K.; Lown, J. W. *Nucleic Acids Res.* **1988**, *16*, 665.

(*Received in USA* 29 *December* 1993; *accepted* 8 *February* 1994)