Biochemical and Biophysical Research Communications **267,** 139–144 (2000)
doi:10.1006/bbrc.1999.1933, available online at http://www.idealibrary.com on IDEAL®

DNA Binding Properties of Novel Distamycin Analogs That Lack the Leading Amide Unit at the N-Terminus

Santanu Bhattacharya*,†,[1] and Mini Thomas*

**Department of Organic Chemistry, Indian Institute of Science, Bangalore 560 012, India; and*
†Chemical Biology Unit, JNCASR, Bangalore 560 012, India

Received November 22, 1999

First examples of distamycin (Dst) analogs which lack hydrogen bond donor or acceptor groups at the N-terminus have been synthesized. The first molecule of this series, which is a bispyrrole peptide, did not exhibit any detectable binding with double-stranded (ds) DNA. However, all other analogs did bind strongly to AT-rich sequences of ds-DNA, with the binding affinities increasing as a function of the number of repeating pyrrole carboxamide units. These results imply that a hydrogen bond donor or acceptor atom per se at the N-terminus is not a prerequisite for DNA binding in the case of pyrrole carboxamide-based Dst analogs. However, in the absence of H-bond donor or acceptor at the N-terminus, a minimum of three pyrrole carboxamide units is necessary for the onset of DNA binding. Beyond this minimum number, the binding affinity increases as a function of the number of pyrrole units, as a result of the greater availability of hydrogen bonding and van der Waals surface. Experiments with poly[d(G-C)] have shown that the presence of the N-terminus formamide group is not inevitable for GC binding of this class of molecules. The observation that the N-terminus formamide unit can be dispensed with suggests that these molecules, which are much easier to synthesize and functionalize, can be used in place of the conventional analogs of distamycin for the development of novel minor groove binders with extended sequence recognition properties.

Key Words: **novel distamycin analogs; absence of leading amide; minimum size requirement; GC binding.**

The minor groove of double helical DNA is the site of noncovalent interactions for a large number of anticancer drugs, antibiotics and antiviral agents (1) which are believed to exert their biological activity by competing with transcription factors or architectural proteins such as E2F (2), TATA-box binding protein (TBP) (3) or DNA topoisomerase I/II (4, 5). The molecular basis of the DNA recognition properties (6) of a number of molecules in this family, notably distamycin (Dst) (7), netropsin (Nt) (8), and Hoechst-33258 (9) has been studied extensively. The AT sequence preference in the binding of these molecules is a consequence of, in particular (i) the sequence-dependent narrow width of the minor groove of the *B*-form of DNA, resulting in stabilization of the complex by van der Waals interaction with the walls of the groove, and (ii) their ability to form specific hydrogen bonds with the donor and acceptor atoms on the minor-groove side of AT-base pairs (6–9). Even though hydrogen bonding and electrostatic interactions contribute to the free energy of binding (10), the information reading process is carried out largely by van der Waals contacts between the ligand and the surface of the minor groove (6). Based on this rationale, design of a number of molecules have been reported which include imidazole (Im), pyridine (Pyn) (11, 12), thiazole (13) or benzene (14) based amides as the basic unit of potentially DNA binding oligopeptides and are termed "lexitropsins" or information reading molecules (13). In these designs, the introduction of heterocyclic rings containing hydrogen bond acceptors such as nitrogen as in Im or Pyn, was based on the realization that these atoms could act as hydrogen bond acceptors for the NH_2 group of guanine in the minor groove and therefore oligo peptides containing such unnatural amino acids would be able to recognize GC rich sequences of DNA. For eventual use in precise genetic targeting, it is necessary to have groove binding ligands, which are capable of discriminating not only AT vs GC base pairs, but also end-for-end reversals of AT and GC base pairs. Such discrimination was realized by designing hairpin polyamides containing Im and or Py rings (15). Based on these studies a set of pairing rules were developed, according to which two amide units are required to recognize one base pair of

[1] To whom correspondence should be addressed at Department of Organic Chemistry, Indian Institute of Science, Bangalore 560 012, India. Fax: +91-80-334-1683; +91-80-344-3529. E-mail: sb@orgchem.iisc.ernet.in.

0006-291X/00 $35.00

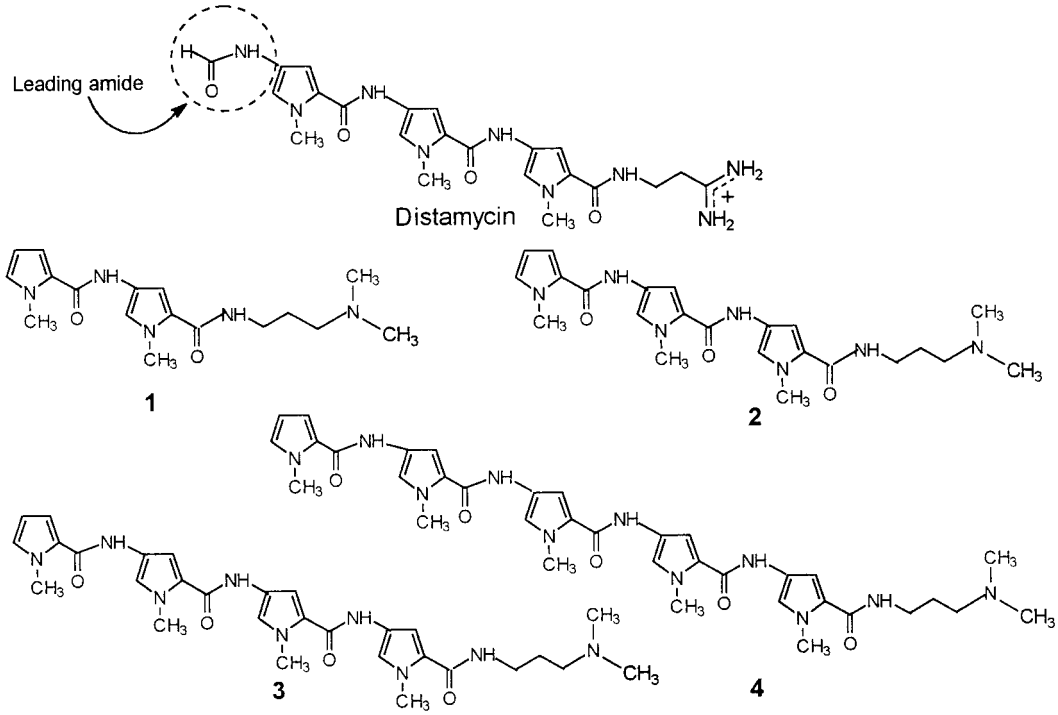

FIG. 1. Chemical structures of distamycin and the newly developed analogs of distamycin that lack the leading amide unit.

DNA and an Im/Py pair targets a GC base pair whereas a Py/Im pair targets a CG pair. A Py/Py combination would be degenerate for AT and TA base pairs (15). Since there is no difference between AT and TA pairs in terms of the availability hydrogen bonding groups in the minor groove (15, 16) such a discrimination in this case could be achieved only on the basis of steric factors and heterodimer hairpins having pyrrole units on one strand and 3-hydroxy pyrrole units on the opposite strand have been demonstrated to be capable of such discrimination (17). It is highly encouraging to note that some of the compounds developed on the basis of these principles are capable of specific gene regulation (18).

In the present study we report the DNA binding properties of a few novel N-methyl pyrrole carboxamide analogs of Dst that lack the N-terminal formamide unit [the "leading amide unit" (16)]. These molecules, **1**–**4,** contain two to five pyrrole carboxamide units (Fig. 1). These have been designed for evaluating the following parameters: (a) the importance of the formamide unit at the N-terminus in the specific DNA binding properties of Dst class of minor groove binders; (b) to assess the minimum size requirement of the peptide for DNA binding in the absence of the leading amide unit; (c) to determine the effect of increasing number of Py based amide units on the DNA binding affinity in the absence of the N-terminus formamide unit, and (d) to investigate the importance of the N-terminal formamide on the binding of Dst analogs to GC-rich regions of DNA.

MATERIALS AND METHODS

All the new compounds reported here have been synthesized using solution phase protocols and have been successfully characterized by ^{1}H NMR, IR and electrospray ionization mass spectrometry. DNA binding experiments were carried out using calf thymus DNA (CT DNA), poly(dA)–poly(dT), poly[d(A-T)] and poly[d(G-C)] (Sigma Chem. Co., St. Louis, MO). Poly(dA)–poly(dT), poly[d(A-T)], and poly[d(G-C)] were used as obtained. CT DNA was purified as described earlier (19).

CD spectra were recorded on a JASCO J-500A spectropolarometer and the data were processed by means of a JASCO-DP-501N data processor. The CD values are expressed as molar ellipticity, $[\theta]$ by following the equation $[\theta] = [100 \times \Psi/l \times c]$ deg.cm^2.dmol^{-1}, where Ψ is the observed ellipticity in degrees, l is the path length of the cell in centimeters and c is the concentration of DNA in base molarity. CD titrations were carried out by adding progressively increasing amounts of the respective ligand to the cuvette, keeping the DNA concentration constant.

Fluorescence spectra were recorded on a Hitachi F-4500 fluorescence spectrophotometer. Known concentrations of a given ligand were added in small increments into solutions containing ethidium bromide (ETBr) ds-DNA complex. After each addition, the mixtures were carefully stirred, followed by recording of the corresponding fluorescence emission spectrum (550–700 nm) upon excitation at 525 nm. The changes in the fluorescence emission peak (590 nm) were plotted against the concentrations of each ligand. The concentration of ligand required to obtain 50% quenching of the maximal fluorescence (C_{50} value) was used to calculate the relative binding constant (20, 21).

RESULTS AND DISCUSSION

Mixed Im/Py and Pyn/Py oligopeptides lacking the leading amide unit have been reported in literature

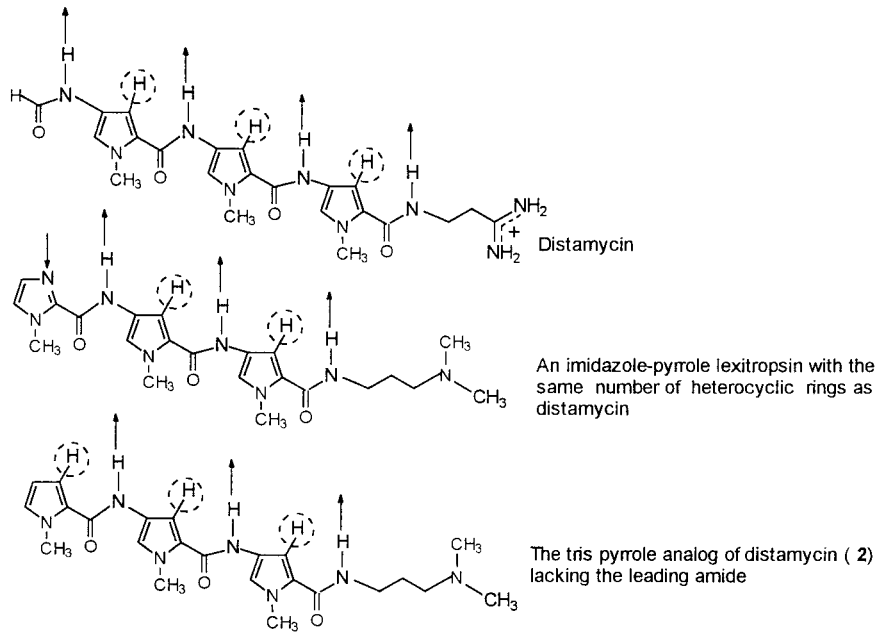

FIG. 2. Comparison of the van der Waals and hydrogen bonding surfaces of distamycin, an imidazole-pyrrole lexitropsin (Refs. 12 and 22) and the newly developed tris-pyrrole analog of distamycin (compound **2**). Hydrogen bond donors are represented by upward arrows and acceptors by downward arrows. Hydrogens involved in van der Waals contact with the minor groove are encircled by dashed circles. It should be noted that in the Im/Py lexitropsin the N-terminus C(3)H has been replaced by a N, as a result of which the van der Waals surface is reduced whereas the number of hydrogen bonding groups remain the same (four) except that the N-terminus donor has been replaced by an acceptor. In the case of **2,** the van der Waals surface remain the same, whereas the number of hydrogen bonding groups has reduced to three.

(11, 22). However, it is important to note that these molecules possess a nitrogen in the N-terminal heterocyclic ring [e.g., N (3) in the case of imidazole] which act as a hydrogen bond acceptor for guanine NH_2 and therefore the number of hydrogen bonding groups remain the same for Dst and such a molecule having three heterocyclic rings (Fig. 2). Thus four synthetic analogs of Dst described herein (**1**–**4**) lack the leading amide unit at the N-terminus and these represent the first examples of Dst analogs that possess neither a hydrogen bond donor nor an acceptor group at the N-terminus.

The DNA binding studies were first carried out using the AT-rich sequences, namely CT DNA (~52% AT) and poly[d(A.T)] (100% AT). Neither the free polyamides, nor the DNA duplexes used, exhibit any CD signals in the ligand absorption region. However, upon addition of ligands **2**–**4,** to all the three forms of DNA used, substantial induced CD signals (ICD) arise in the 300–380 region. Since this induced cotton effect is outside the CD spectrum of DNA, it directly reflects the environment of the bound ligand molecules. The CD signals observed in this case were characterized by two ICD maxima with +ve and −ve amplitudes respectively in the wave length region centered around 320 and 280 nm, as exemplified by the titration of **3** with CT DNA shown in Fig. 3. The magnitude of the observed ICD signal (normalized for DNA concentration) increased with the number of Py units in the oligopeptide (Fig. 4A). Apparently, the number of chromophoric pyrrole carboxamide residues contributes to the intensity and location of the negative and positive ICD bands. The magnitude of ICD signal obtained for **2** with poly[d(A.T)] was ~2 times higher compared to that of CT DNA, whereas in the case of **3** and **4** it was higher by about 20%.

The observed ICD signals saturated in all the cases mentioned above at a certain compound to nucleotide phosphate ([*D*]/[*P*]) ratio characteristic of each ligand-duplex type. The general trend was a decrease in the saturation value of [*D*]/[*P*] with the increase in the number of Py rings in the peptide sequence, as exemplified by the titration of **2**–**4** with poly[d(A.T)] in Fig. 4A, suggesting an increase in the binding site size with the length of the peptide as expected (22, 23). To a first approximation, the magnitude of the ICD signal can be taken as an indication of the extent of binding of a given ligand to a certain type of duplex. Therefore the saturation in ICD can be used for estimating the binding site size of the ligand (the number of base pairs covered by one ligand molecule) for a given type of duplex, as demonstrated for **4** and poly[d(A.T)] in Fig. 4B. The same procedure was used for the estimation of the binding site size for **2** and **3** with this duplex and were conforming with a head-to-tail dimeric binding mode (Fig. 5), as was observed for Dst (24) and most of the analogs as well (12, 22). The observed value of 8.9 in the case of **4** is higher than the calculated value of

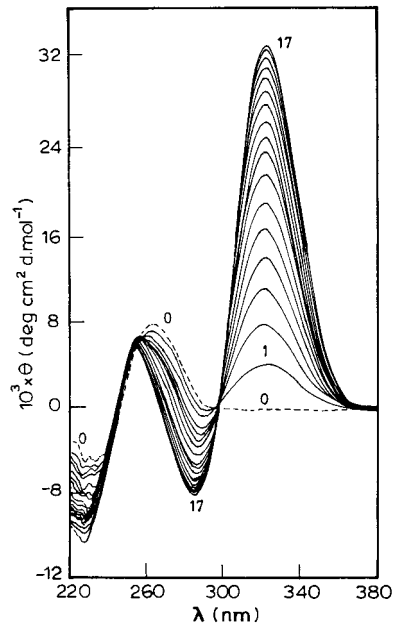

FIG. 3. Circular dichroism spectra of **3**/CT DNA complexes at several [ligand] to [nucleotide] ratios ([*D*]/[*P*] ratios) in 10 mM Tris–HCl buffer (pH 7.4) containing 40 mM NaCl at 25°C. Trace "0" refers to the CD spectrum of free CT DNA in the same buffer. Traces 1 and 17 correspond to the complexes at [*D*]/[*P*] ratios 0.027 and 0.49, respectively.

6.9 (23) and this can be explained by assuming a slightly "slipped" binding mode in this case.

Quite surprisingly, ligand **1** failed to exhibit any detectable ICD under the same conditions. The absence of any detectable ICD in this case is indicative of the absence of interaction between this ligand and DNA. This observation is important and suggests that a minimum of three pyrrole carboxamide units are necessary for the onset of DNA binding in the case of "all pyrrole" polyamides lacking the leading amide unit. This essentially would mean that a minimum of three hydrogen bonds are necessary for holding a Dst class of DNA binding compound possessing a single positive charge in the minor groove.

The well-known ethidium bromide (ETBr) displacement assay was then employed to estimate the apparent binding constants of these ligands to ds-DNA. This assay is based on the displacement of intercalatively bound ETBr from the duplex DNA upon addition of another DNA-binding ligand. Apparent Binding constants (K_{app}) of such ligands to ds-DNA can be estimated and compared by measuring the loss of ETBr fluorescence as a function of added ligand. The K_{app} values were calculated from K_{ETBr} [ETBr] = K_{app}.[Ligand], where [ETBr] and K_{ETBr} are the concentrations and binding constants of ETBr respectively and [Ligand] is the concentration of ligand at 50% of maximal ETBr fluorescence. The binding constant of ETBr was taken to be 1×10^7 M^{-1} (20). The apparent binding constants for the three oligopeptides, **2**–**4** calculated this way are presented in Table 1. It is evident that the binding constants increase for all the three forms of DNA as the number of pyrrole carboxamide unit increases. Moreover, for a given oligopeptide, the binding constant is higher for poly[d(A.T)] and poly[dA]–poly(dT) compared to that of CT DNA (~48% GC), suggesting that these molecules retain their AT specificity despite the absence of the leading amide unit. It should also be noted that compound **4** show higher binding affinities compared to Dst. This observation should be analyzed in the light of the fact that in comparison with Dst, compound **3** carries one extra hydrogen for van der Waals interaction (C(3)H of the Py at the N-terminus, Fig. 2) and compound **4** carries one extra amide NH and a C(3)H in addition to the terminal Py C(3)H. To test the possibility of these ligands binding to poly[d(G.C)] sequences, **3** was chosen as a representative example (This molecule possesses equal number of carbonyl groups as distamycin except that the N-terminus formamide is absent). These experiments assume significance considering the fact that Dst is known to exhibit ICD signals with poly[d(G.C)] and the formamide group at the

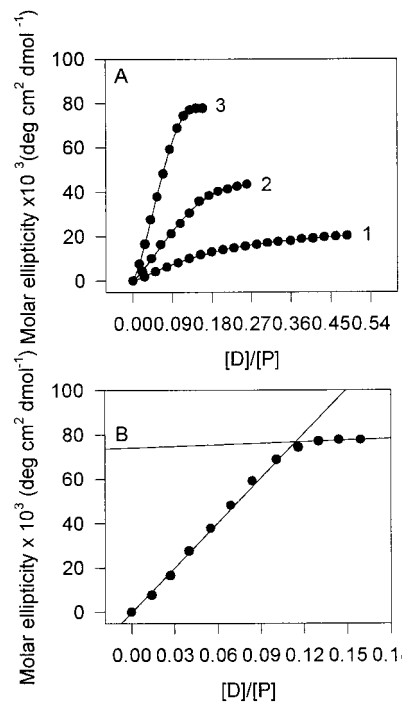

FIG. 4. (A) Molar ellipticities vs [*D*]/[*P*] ratio plots for the CD titration of compounds **2**–**4** with poly[d(A.T)]. Titrations were done in 10 mM Tris–HCl buffer (pH 7.4) containing 40 mM NaCl. Plots 1–3 represent the plots for ligands **2**–**4,** respectively. (B) Calculation of saturation [*D*]/[*P*] and hence the binding site size from the above plot for **4.** Considering a head-to-tail binding model, the binding site size calculated this way for **2**–**4** were 4.8, 5.9, and 8.9, respectively. These values were close to the expected values of 4.6, 5.8, and 6.9 (Refs. 22 and 23), respectively, for compounds **2**–**4.** See text for details.

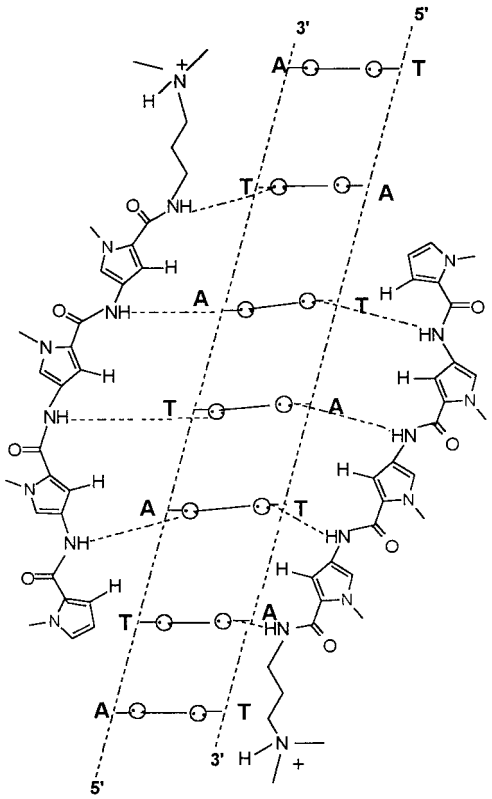

FIG. 5. Schematic representation of the proposed 2:1 anti parallel side-by-side binding model for the complex formed between 3 and a stretch of poly[d(A.T)]. Circles with double dots represent lone pairs of N-3 of adenines and O-2 of thymines. Dashed lines connecting the NHs and the lone pairs represent putative hydrogen bonds.

N-terminus was also implicated in the formation of such complexes (23). Interestingly, intense ICD signals could be observed upon progressive addition of **3** to a solution of poly[d(G.C)] (not shown). Based on the experiments using m7G-DNA (in which the guanine NH_2 is methylated) and deformylated distamycin (Dst-D, in which the N-terminal NH_2 group is free), Luck *et al.* had suggested the formation of hydrogen bonds between the amino group of the guanine bases and the C═O group of the formamide terminus (in addition to the other C═O groups of the amide units) as the driving force for the complex formation between this class of compounds and GC sequences (23). In their experiments, Dst-D exhibited practically no binding with poly(dG)–poly(dC) in 20 mM NaCl whereas in the present case **3** exhibited intense ICD signals with poly-[d(G.C)] even in 40 mM NaCl. These observations suggest that a minimum of four CO groups are necessary for the maintenance of GC binding. The N-terminal formamide carbony per se is not a prerequisite for this type of binding.

To ascertain the nature of the interaction between poly[d(G.C)] and **3,** effect of addition of salt (NaCl) on the ICD signal was monitored by adding increasing amounts of NaCl to a preformed complex of **3** and poly[d(G.C)] (10 mM Tris–HCl, pH 7.4, 40 mM NaCl). The observed ICD signals collapsed almost completely (to about 2% of the original intensity) when the salt concentration was increased to 200 mM (not shown). This indicates the highly electrostatic nature of the complex. Compound **3** and the AT-rich sequences of DNA on the other hand exhibited substantial ICD signals even in the presence of ~4.8 M NaCl, confirming the inference that the later described complexes are due to specific interactions between the ligand and DNA.

In conclusion, we have been able to demonstrate that hydrogen bond donor or acceptor groups in the N-terminus per se is not a prerequisite for achieving DNA binding in the case of pyrrole carboxamide based Dst analogs. However, in the absence of N-terminus hydrogen bond donor or acceptor, a minimum of three pyrrole carboxamide units are necessary for the onset of DNA binding. In a general sense this essentially would mean that a minimum of three hydrogen bonds are necessary for a Dst class of DNA binding compound possessing a single positive charge to be able to bind in the minor groove. Beyond this minimum number, in the examples studied here, the binding affinity increases as a function of the number of Py units, as a result of the greater availability of hydrogen bonding and van der Waals surface. Experiments with GC-rich

TABLE 1

Apparent Binding Constants of the Oligopeptides for the Three Types of AT-Rich Sequences of DNA Used

	Poly(dA)–poly(dT)		Poly[d(A.T)]		CT DNA	
Compound	C_{50}[a] (μM)	$10^6 \times K_{app}$ (M^{-1})	C_{50} (μM)	$10^6 \times K_{app}$ (M^{-1})	C_{50} (μM)	$10^6 \times K_{app}$ (M^{-1})
2	9.0	1.6	37.9	0.42	275.0	0.06
3	0.23	63	0.7	33.0	14.44	101.0
4	0.16	92	1.23	12.0	3.23	4.52
Distamycin[b]				34.8		77.0

[a] C_{50} is defined as the amount of pyrrole oligopeptide required for 50% inhibition of ethidium fluorescence.
[b] Taken from Ref. 21.

sequences clearly demonstrated that the N-terminal formamide group is not a prerequisite for GC binding.

Unlike Dst and Nt which are insoluble in organic solvents, all the compounds reported here are soluble in organic solvents in their free base form (the tertiary amino group at the C-terminus gets protonated in water and thus imparts positive charge to the molecule). Hence, these could easily be functionalized by suitable reactions in common organic solvents. Moreover, the fact that the N-terminus formamide group can be dispensed with point to the possibility of using these compounds, which are considerably easier to synthesize and functionalize, in place of the conventional analogues of Dst that retain a formamido or acetamido group at the N-terminus. Currently we are developing "affinity cleavage agents" based on these templates by the attachment of appropriate redox-active (25, 26) or photoactive (27) groups. Such compounds should eventually find applications as "nonprotein based restriction enzymes" that can cleave specific sequences of DNA, depending on the choice of the groove binding moiety.

REFERENCES

1. Dervan, P. B. (1986) *Science* **232,** 464–471.
2. Welch, J. J, Rauscher, F. J., III, and Beerman, T. A. (1994) *J. Biol. Chem.* **269,** 31051–31058.
3. Chiang, S.-Y., Welch, J., Rauscher, F. J., III, and Beerman, T. A. (1994) *Biochemistry* **33,** 7033–7040.
4. Chen, A. Y., Yu, C., Gatto, B., and Liu, L. F. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8131–8135.
5. Simon, H., Wittig, B., and Zimmer, C. (1994) *FEBS Lett.* **353,** 79–83.
6. Iida, H., Jia, G., and Lown, J. W. (1999) *Curr. Opin. Biotechnol.* **10,** 29–33.
7. Coli, M., Frederik, C. A, Wang, H.-J., and Rich, A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8385–8389.
8. Coll, M., Aymami, J., Gigis, A., van der Marel, G. A., van Boom, J. H., Rich, A., and Wang, A. H.-J. (1989) *Proc. Natl. Acad. Sci. USA* **28,** 310–320.
9. Quintana, J. R., Lippanov, A. A., and Dickerson, R. E. (1991) *Biochemistry* **30,** 10294–10306.
10. Marky, L. A., Curry, J., and Breslauer, K. J. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4359–4363.
11. Lown, J. W., Krowicki, K., and Bhat, U. G. (1986) *Biochemistry* **25,** 7408–7416.
12. Wade, W. S., Mrksich, M., and Dervan, P. B. (1992) *J. Am. Chem. Soc.* **114,** 8783–8794.
13. Rao, K. E., Bathni, Y., and Lown, J. W. (1990) *J. Org. Chem.* **55,** 728–737.
14. Yan, Y., and Gong, B. (1997) *Biochem. Biophys. Res. Commun.* **240,** 557–560.
15. Wemmer, D. E., and Dervan, P. B. (1997) *Curr. Opin. Struct. Biol.* **7,** 355–361.
16. Kopka, M. L., Goodswell, D. S., Han, G. W., Chiu, T. K., and Lown, J. W. (1997) *Structure* **5,** 1033–1046.
17. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. (1998) *Nature* **391,** 468–471.
18. Gottesfield, L., Neely, L., Trauger, J. W., Baird, E. E., and Dervan, P. B. (1997) *Nature* **387,** 202–205.
19. Bhattacharya, S., and Thomas, M. (1998) *J. Ind. Chem. Soc.* **75,** 716–724.
20. Morgan, A. R., Lee, J. S., Pullyblank, D. E., Murray, N. L., and Evans, D. H. (1979) *Nucleic Acids Res.* **4,** 547–569.
21. Lee, M., Rhodes, A. L., Wyatt, M. D., Forrow, S., and Hartly, J. H. (1995) *Biochemistry* **32,** 4237–4245.
22. Kelly, J. J., Baird, E. E., and Dervan, P. B. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 6981–6985.
23. Luck, G., Zimmer, C., Rinert, K. E., and Arcamone, F. (1977) *Nucleic Acids Res.,* **4,** 2655–2670.
24. Fagan, P., and Wemmer, D. E. (1992) **114,** 1080–1081.
25. Bhattacharya, S., and Mandal, S. S. (1995) *Chem. Commun.,* 2489–2490.
26. Mandal, S., Varshney, U., and Bhattacharya, S. (1997) *Bioconjugate Chem.* **8,** 798–812.
27. Bhattacharya, S., and Mandal, S. S. (1996) *Chem. Commun.,* 1515–1516.