Nonnatural Deoxyribonucleoside D_3 Incorporated in an Intramolecular DNA Triplex Binds Sequence-Specifically by Intercalation

Karl M. Koshlap,† Paul Gillespie,‡ Peter B. Dervan,*,‡ and Juli Feigon*,†

Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California Los Angeles, California 90024
Division of Chemistry and Chemical Engineering California Institute of Technology Pasadena, California 91125

Received May 26, 1993

Oligonucleotide-directed triple helix formation is one of the most powerful methods for the sequence-specific recognition of double-helical DNA.[1,2] Pyrimidine oligonucleotides bind purine tracts in the major groove of DNA parallel to the purine Watson–Crick strand through the formation of specific Hoogsteen-type hydrogen bonds.[1-4] Specificity is derived from thymine (T) recognition of adenine·thymine (A·T) base pairs (T·A·T triplet) and N3-protonated cytosine (C+) recognition of guanine–cytosine (G·C) base pairs (C + G·C triplets).[1-4] The sequence-specific recognition of double-helical DNA by a third strand to form a triple helix is limited to mostly purine tracts. Although G in the third strand has been found to specifically bind to T·A,[5] the lower stability of the G·T·A triplet and its dependence on the sequence of the neighboring triplets reveals that this will have limitations.[5-8] In an attempt to extend the recognition code to all four Watson-Crick base pairs, the nonnatural deoxyribonucleoside 1-(2-deoxy-β-D-ribofuranosyl)-4-(3-benzamido)phenylimidazole [D_3][9] was synthesized and incorporated into pyrimidine DNA oligonucleotides (Figure 1a). It was found that D_3 selectively recognizes both T·A and C·G Watson-Crick base pairs within the pyrimidine·purine·pyrimidine triple-helix motif.[9] This was also found to have a nearest neighbor dependence.[5b]

In order to determine the binding mode of the nonnatural deoxyribonucleoside D_3 with the Watson-Crick base pair T·A,

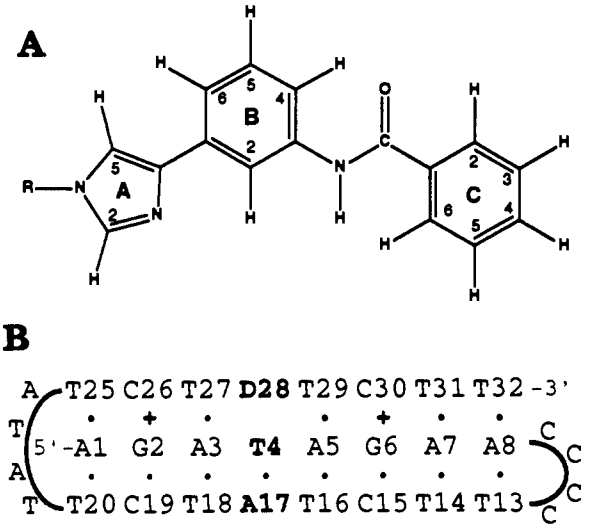

Figure 1. (A) Chemical structure of the nonnatural deoxyribonucleoside D_3 with ring labeling and numbering scheme used. R is deoxyribose. (B) Schematic of the proposed folding of the 32-base intramolecular triplex with base numbering scheme used. Hydrogen bonds are indicated by · except for those of protonated C, indicated by (+).

the oligonucleotide with the sequence d(AGATAGAACCCC-TTCTATCTTATATCTD_3TCTT) was synthesized as previously described[9] and studied by one- and two-dimensional ^{1}H NMR spectroscopy. The molecule was designed to fold into an intramolecular triplex, similar to those we have previously reported[8,10,11] (Figure 1b). NMR spectra of this DNA oligonucleotide (100 mM NaCl, 5 mM $MgCl_2$, pH 5.2, 2 mM strand) indicate that the molecule forms an intramolecular triplex. Fifteen hydrogen-bonded imino proton resonances are observed in the one-dimensional ^{1}H spectra, and NOESY[12] spectra in H_2O (spectra not shown) were used to assign these to the eight Watson-Crick and seven Hoogsteen base pairs indicated in Figure 1b.[13] Downfield-shifted amino proton resonances, characteristic of protonated cytosine bases, and NOE cross peaks between imino protons and sugar resonances of the Hoogsteen base-paired strand are also observed, as expected.[8,10,11]

Although the NMR spectra indicate that the D_3-containing oligonucleotide forms an intramolecular triplex, there are some unexpectedly missing cross peaks in the NOESY spectra. Sequential connectivities observed in NOESY spectra in D_2O[14] from base–H1′ (and base–H_2′,H2″) along each strand of the triplex are missing between T4H1′ and A5H8, between T16H1′ and A17H8, and between T27H1′ and $D_3$28. There is also a break in the imino–imino sequential connectivities between T4·A17 and A5·T16. In addition, no NOE cross peaks are observed between the base protons of D_3 and the T27·A3·T18 triplet.

* Authors to whom inquiries should be directed.
† University of California.
‡ California Institute of Technology.

(1) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645–650.
(2) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J. L.; Thuong, N. T.; Lhomme, J.; Hélène, C. *Nucleic Acids Res.* **1987**, *15*, 7749–7760.
(3) (a) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637–640. (b) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859–7870. (c) de los Santos, C.; Rosen, M.; Patel, D. *Biochemistry* **1989**, *28*, 7282–7289.
(4) For early studies on triple-stranded polynucleotides, see: (a) Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023–2024. (b) Lipsett, M. N. *Biochem. Biophys. Res. Commun.* **1963**, *11*, 224–228. (c) Lipsett, M. N. *J. Biol. Chem.* **1964**, *239*, 1256–1260. (d) Michelson, A. M.; Massoulié, J. Guschlbauer, W. *Prog. Nucleic Acid Res. Mol. Biol.* **1967**, *6*, 83-141. (e) Felsenfeld, G.; Miles, H. T. *Annu. Rev. Biochem.* **1967**, *36*, 407–448. (f) Howard, F. B.; Frazier, J.; Lipsett, M. N.; Miles, H. T. *Biochem. Biophys. Res. Commun.* **1964**, *17*, 93–102. (g) Miller, J. H.; Sobell, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 1201–1205. (h) Morgan, A. R.; Wells, R. D. *J. Mol. Biol.* **1968**, *37*, 63–80. (i) Lee, J. S.; Johnson, D. A.; Morgan, A. R. *Nucleic Acids Res.* **1979**, *6*, 3073–3091.
(5) (a) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967–971. (b) Kiessling, L. L.; Griffin, L. C.; Dervan, P. B. *Biochemistry* **1992**, *31*, 2829–2834.
(6) (a) Belotserkovskii, B. P.; Veselkov, A. G.; Filippov, S. A.; Dobrynin, V. N.; Mirkin, S. M.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **1990**, *18*, 6621–6624. (b) Roberts, R. W.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 9397–9401. (c) Mergny, J.-L.; Sun, J.-S.; Rougee, M.; Montenay-Garestier, T.; Barcelo, F.; Chomilier, J.; Hélène, C. *Biochemistry* **1991**, *30*, 9791–9798.
(7) Radhakrishnan, I.; Gao, X.; de los Santos, C.; Live, D.; Patel, D. J. *Biochemistry* **1991**, *30*, 9022–9030.
(8) Wang, E.; Malek, S.; Feigon, J. *Biochemistry* **1992**, *31*, 4838–4846.
(9) Griffin, L. C.; Kiessling, L. L.; Beal, P. A.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 7976–7982.
(10) (a) Sklenář, V.; Feigon, J. *Nature* **1990**, *345*, 836–838. (b) Macaya, R.; Wang, E.; Schultze, P.; Sklenář, V.; Feigon, J. *J. Mol. Biol.* **1992**, *225*, 755–773.
(11) Macaya, R. F.; Gilbert, D. E.; Malek, S.; Sinsheimer, J.; Feigon, J. *Science* **1991**, *254*, 270–274.
(12) (a) Macura, S.; Ernst, R. R. *Mol. Phys.* **1980**, *41*, 95–117. (b) Kumar, A.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1–6.
(13) Sequential assignment of the imino proton resonances of the Watson-Crick and Hoogsteen base-paired strands, respectively, were obtained from NOESY spectra in H_2O as previously described (refs 10, 11).
(14) Sequential assignments of the natural base and the sugar proton resonances of the triplex were obtained essentially as described (ref 10b). The D_3 base protons were assigned by a combination of NOESY, P.COSY (Mueller, L. *J. Magn. Reson.* **1987**, *72*, 191–196. Marion, D.; Bax, A. *J. Magn. Reson.* **1988**, *80*, 528–533), HOHAHA (Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521–528. Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *65*, 355–360) and proton–carbon HMQC (Bax, A.; Griffey, R. H.; Hawkins, B. L. *J. Magn. Reson.* **1983**, *55*, 301–315) experiments. The assignments will be presented elsewhere (ref 16).

0002-7863/93/1515-7908$04.00/0

Table I. Internucleotide NOEs of $D_3$28[a]

	AH2	AH5	BH2	BH4	BH5	BH6	CH2/6	CH3/5	CH4
T4H1′				w	w				
T4H3′					w	w			
T4H2″				m	s	w			
T4H2′				m*	s	w			
T4H6					m	m			
T4Me	w	m*	s			m			
A17H1′								m*	
A17H8								m	m
A17H2							m	w	
T29H1′		m				w			
T29H6	w	w							
T29Me	s	w							
T29imino					m	m			
A5H8				m*	s	m			
A5H2								w	
T16H1′								m*	w
T16H3′								w	w
T16H2″								s	s
T16H2′								s	s
T16H6								m	m
T16Me							m	s	m
T16imino								m	

[a] NOEs are classified as strong (s), medium (m), or weak (w) intensity. An asterisk indicates some ambiguity in NOE intensity due to cross peak overlap. NOE cross peaks were obtained from a NOESY spectrum which was acquired essentially as previously described,[15] with τ_m = 200 ms at 25 °C. The data for the imino NOEs are from a NOESY spectrum in H_2O, with τ_m = 100 ms at 1 °C.

The NOE cross peaks observed between the D_3 base and other nucleotides in the triplex are summarized in Table I. Several cross peaks are observed between the C ring protons and T16 and A17 and between the B ring protons and T4 and A5. Cross peaks from the A ring protons are observed to T29 but not to T27. These NOE cross peaks are only consistent with $D_3$28 being positioned *between* the T29·A5·T16 triplet and a T4·A17 base pair, rather than forming a $D_3$28·T4·A17 triplet, as illustrated in Figure 2. Particularly compelling evidence for binding of D_3 by intercalation are the NOE cross peaks between BH4, BH5, BH6 and T4H2′, T4 H2″, between BH4, BH5, BH6 and A5H8, between CH2/6, CH3/5, and A17H2, and between CH3/5 and A5H2. The intercalation of $D_3$28 between T4·A17 and T29·A5·T16 is consistent with the breaks in sequential connectivities discussed above. Similar breaks in sequential connectivities are observed in intercalative drug–DNA complexes.[15] Intercalation of D_3 also provides an explanation for the large upfield shifts of the T4H6 (6.32 ppm) and T4Me (0.34 ppm) resonances, which are >0.9 ppm upfield of the other TH6 and TMe resonances in the triplex, respectively, due to ring current shifts from the stacked phenylimidazole moiety of D_3.

(15) For example, see: (a) Gilbert, D. E.; Feigon, J. *Biochemistry* **1991**, *30*, 2483–2494. (b) Addess, K. J.; Gilbert, D. E.; Olsen, R. K.; Feigon, J. *Biochemistry* **1992**, *31*, 339–350.

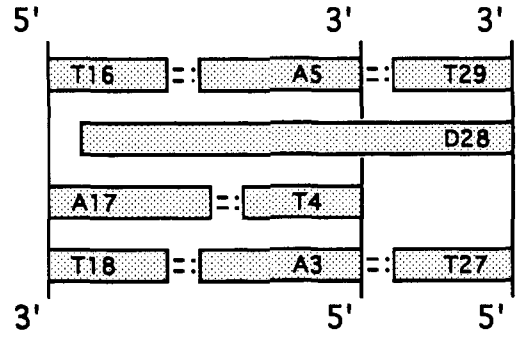

Figure 2. Schematic illustrating the binding mode of D_3 and neighboring nucleotides.

The intercalative binding mode of D_3 shown here provides a rationale for the results of previous studies on the sequence specificity and nearest-neighbor effects of triplexes containing the D_3 nucleotide in the Hoogsteen-paired strand.[5b,9] Those studies indicated that in a triplex containing a D_3TA "triplet", only the 3′ neighboring triplet (i.e., 3′ along the purine strand) affected the triplex stability. This is to be expected, since intercalation of D_3 places it next to a T·A base pair rather than the 5′ neighboring triplet. The best intercalation sites for D_3 appear to be TpA and CpA, followed by ApA and then GpA; there appears to be little binding at NpG sites. Thus, D_3 shows sequence-specific intercalation at YpA sites. Triplexes containing D_3 opposite T·A and C·G have comparable stabilities, and NMR studies on a related D_3CG triplex indicate that the binding mode is the same.[16]

Covalent attachment of intercalating agents at the 5′ end of a third-strand oligonucleotide has previously been shown to increase the stability of DNA triplexes.[17] The binding mode of D_3 is unique in that it not only binds by intercalation but skips a potential base pair to do so. Thus, non-purine bases can be accommodated in the purine strand of DNA triplexes in an entirely new way. This unexpected result opens the door for a new design approach to achieving sequence specificity and stability in mixed sequence DNA triplexes.

Acknowledgment. This work was supported by grants from NIH (R01 GM 37254-01) and NSF Presidential Young Investigator Award (DMB 89-58280), with matching funds from AmGem Inc., DuPont/Merck Pharmaceuticals, Monsanto Co., and Sterling Drug Inc. to J.F., and Office of Naval Research grant to P.B.D.

(16) Koshlap, K. M.; Wang, E.; Gillespie, P.; Dervan, P. B.; and Feigon, J., manuscript in preparation.

(17) (a) Sun, J.-S.; François, J.-C.; Montenay-Garestier, T.; Saison-Behmoaras, T.; Roig, V.; Thuong, N. T.; Hélène, C. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9198–9202. (b) Grigoriev, M.; Praseuth, D.; Robin, P.; Hemar, A.; Saison-Behmoaras, T.; Dautry-Varsat, A.; Thuong, N. T.; Hélène, C.; Harel-Bellant, A. *J. Biol. Chem.* **1992**, *267*, 3389–3395.