

## Toward a Designed, Functioning Genetic System with Expanded-Size Base Pairs: Solution Structure of the Eight-Base xDNA Double Helix

Stephen R. Lynch, Haibo Liu, Jianmin Gao, and Eric T. Kool\*

Contribution from the Department of Chemistry, Stanford University,  
Stanford, California 94305-5080

Received August 2, 2006; E-mail: kool@stanford.edu

**Abstract:** We describe the NMR-derived solution structure of the double-helical form of a designed eight-base genetic pairing system, termed xDNA. The benzo-homologous xDNA design contains base pairs that are wider than natural DNA pairs by ca. 2.4 Å (the width of a benzene ring). The eight component bases of this xDNA helix are A, C, G, T, xA, xT, xC, and xG. The structure was solved in aqueous buffer using 1D and 2D NMR methods combined with restrained molecular dynamics. The data show that the decamer duplex is right-handed and antiparallel, and hydrogen-bonded in a way analogous to that of Watson–Crick DNA. The sugar–phosphate backbone adopts a regular conformation similar to that of B-form DNA, with small dihedral adjustments due to the larger circumference of the helix. The grooves are much wider and more shallow than those of B-form DNA, and the helix turn is slower, with ca. 12 base pairs per 360° turn. There is an extensive intra- and interstrand base stacking surface area, providing an explanation for the greater stability of xDNA relative to natural DNA. There is also evidence for greater motion in this structure compared to a previous two-base-expanded helix; possible chemical and structural reasons for this are discussed. The results confirm paired self-assembly of the designed xDNA system. This suggests the possibility that other genetic system structures besides the natural one might be functional in encoding information and transferring it to new complementary strands.

### Introduction

We have recently undertaken a program to design and develop a new functioning genetic system using expanded-size base pairing. The steps to this end are to design new nucleobases and new nucleoside building blocks, to assemble a new DNA-like molecule that can encode sequence information, and ultimately to develop enzymes that can replicate this genetic set, establishing storage and transfer of genetic information. This strategic program serves two purposes. First, by recapitulating the properties and activities of the natural genetic system, we hope to gain new insights into how biological information is

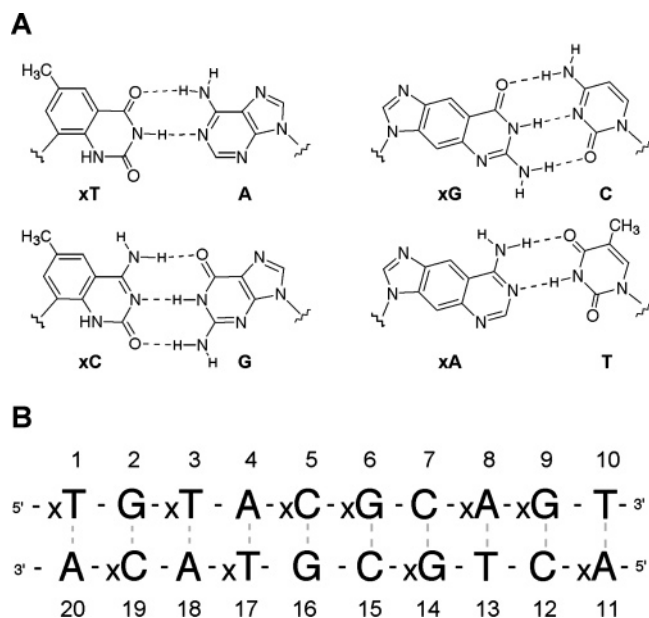
stored and transferred in the natural system. Second, we expect that both the nucleotides and oligomers of xDNA, which are inherently fluorescent, may serve as useful biophysical and biomedical tools for probing living systems.

A number of laboratories have been designing new bases and base pairs to function in the context of the natural genetic system.<sup>1</sup> The present approach is different from that previous work because the expanded pairing set has not just one but all base pairs replaced, and although structurally related, these pairs have a geometry different from that of Watson and Crick. Because of the large variations in base size, eight-base xDNA strands are not able to bind natural DNA and are thus structurally orthogonal to the natural genetic system. In addition to studying modified nucleobases, a few laboratories have also investigated the replication of modified nucleic acid backbones.<sup>2</sup> Our own research has focused instead on the bases because they (as opposed to the deoxyribose–phosphodiester backbone) are the site of information storage by a defining sequence.

Our molecular design for expansion of pairs involves benzo homologation of DNA bases,<sup>3</sup> a strategy first reported by Leonard in a *lin*-benzoadenine ribonucleotide in 1977.<sup>4</sup> When

(1) (a) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1993**, *32*, 10489–10496. (b) Rappaport, H. P. *Nucleic Acids Res.* **1988**, *16*, 7253–7267. (c) Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863–1872. (d) Tae, E. L.; Wu, Y.; Xia, G.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123*, 7439–7440. (e) Mitsui, T.; Kitamura, A.; Kimoto, M.; To, T.; Sato, A.; Hirao, I.; Yokoyama, S. *J. Am. Chem. Soc.* **2003**, *125*, 5298–5307. (f) Matray, T. J.; Kool, E. T. *J. Am. Chem. Soc.* **1998**, *120*, 6191–6192. (g) Henry, A. A.; Romesberg, F. E. *Curr. Opin. Chem. Biol.* **2003**, *7*, 727–733. (h) Lai, J.; Qu, J.; Kool, E. T. *J. Am. Chem. Soc.* **2004**, *126*, 3040–3041. (i) Johnson, S. C.; Sherrill, C. B.; Marshall, D. J.; Moser, M. J.; Prudent, J. R. *Nucleic Acids Res.* **2004**, *32*, 1937–1941. (j) Atwell, S.; Meggers, E.; Spraggon, G.; Schultz, P. G. *J. Am. Chem. Soc.* **2001**, *123*, 12364–12367. (k) Tanaka, K.; Yamada, Y.; Shionoya, M. *J. Am. Chem. Soc.* **2002**, *124*, 8802–8803. (l) Weizman, H.; Tor, Y. *J. Am. Chem. Soc.* **2001**, *123*, 3375–3376. (m) Zhang, L.; Meggers, E. *J. Am. Chem. Soc.* **2005**, *127*, 74–75. (n) Sismour, A. M.; Benner, S. A. *Nucleic Acids Res.* **2005**, *33*, 5640–5646. (o) Minakawa, N.; Kojima, N.; Hikishima, S.; Sasaki, T.; Kiyosue, A.; Atsumi, N.; Ueno, Y.; Matsuda, A. *J. Am. Chem. Soc.* **2003**, *125*, 9970–9982. (p) Switzer, C.; Sinha, S.; Kim, P. H.; Heuberger, B. D. *Angew. Chem., Int. Ed.* **2005**, *44*, 1529–1532. (q) Adelfinskaya, O.; Nashine, V. C.; Bergstrom, D. E.; Davison, V. J. *J. Am. Chem. Soc.* **2005**, *127*, 16000–16001.

(2) (a) Jung, K. H.; Marx, A. *Cell Mol. Life Sci.* **2005**, *62*, 2080–2091. (b) Chaput, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 9274–9275. (c) Eschenmoser, A. *Origin Life Evol. Biosphere* **2004**, *34*, 277–306. (3) (a) Godde, F.; Toulme, J. J.; Moreau, S. *Biochemistry* **1998**, *37*, 13765–13775. (b) Okamoto, A.; Tanaka, K.; Saito, I. *J. Am. Chem. Soc.* **2003**, *125*, 4972–4973. (4) Scopes, D. I.; Barrio, J. R.; Leonard, N. J. *Science* **1977**, *195*, 296–298.



**Figure 1.** Structures and sequences in this study: (A) proposed xDNA base pairs; (B) strand sequences and residue numbering.

extended to new designs for both pyrimidines and purines, this produces size-expanded bases and base pairs.<sup>5</sup> Two geometries for size expansion have been explored: xDNA<sup>5a,6</sup> and yDNA,<sup>7</sup> which have related benzo homologation but with different vector orientations of expansion. The present study focuses on the former design. xDNA helices are more thermally stable than DNA and have a similar sequence selectivity against mismatches.<sup>8</sup> xDNA bases (Figure 1) are fluorescent, and oligomeric strands composed of the four expanded bases can recognize natural DNA and RNA, suggesting possible applications in sensing and imaging of biological sequences.

A recent report described the completion of the synthesis of the four expanded xDNA components and their assembly into eight-component double-stranded complexes.<sup>8</sup> To function like a biologically active genetic system, such polymers presumably must self-assemble into pairs of bases having a regular structure (Figure 1). An early structural study of a two-base-expanded DNA helix<sup>9</sup> revealed an antiparallel paired geometry. However, an eight-base system is much more complex and raises several important questions. For example, are all the pairs in the full xDNA pairing system isoplanar and hydrogen-bonded, analogous to DNA? Are all base pairs, involving expanded pyrimidines as well as expanded purines, structurally isomorphous? Is the helix antiparallel and right-handed, as with most sequences of natural DNA? What is the rise and twist of the helix, and how does this affect the dimensions of the grooves? How is the structure of the full, complex-sequence eight-base xDNA helix different from that of the preliminary two-base system

and from that of DNA itself? The present study addresses these questions. We find that the eight-base xDNA genetic strands do assemble into a paired helix with remarkable similarities to natural DNA, and yet with distinct structural features that arise from the large sizes of the novel base pairs. The results also provide insight into the high stability of the expanded helical system.

## Results

**Sequence Design and Properties.** For structural studies we chose a sequence putatively containing only expanded pairs, i.e., each pair containing one benzo-homologous base and one natural complementary base. Our goal was to evaluate structure in a case of greatest complexity, where all four expanded bases were present (as opposed to an early structure, which contained only the xA base). Ideally, the expanded bases would be scattered on both strands, and a wide variation of nearest neighbor arrangements would be present, to allow for several different stacking arrangements. Because of the large number of nucleotide components (8), xDNA has 64 possible nearest neighbor arrangements, in contrast to DNA, where there are 16.

We chose the decamer sequence shown in Figure 1B to meet these requirements. The sequence contains all eight components of xDNA scattered on both strands, and all nine nearest neighbors are unique. The pair of oligomers having this sequence was recently shown to form a 1:1 complex with a cooperative melting transition at 63 °C,<sup>8</sup> but no other structural information about helicity and pairing was known. We prepared these two oligomers on larger scales for NMR studies and purified them by reversed-phase HPLC.

**Acidity of dxG.** Preliminary 1D NMR studies in water (see below) suggested that the imino proton of dxG was in more rapid exchange with solvent than is commonly seen for dG (Figure 2). We therefore explored the effect of benzo homologation on the acidity of this nucleobase by measuring the acidic  $pK_a$  of the free nucleoside in water. This was done following a published protocol,<sup>10</sup> evaluating perturbations in pH with small equivalents of base added. The data yielded a  $pK_a$  value of  $7.3 \pm 0.3$  for dxG. This base-induced ionization, presumably of the imino N–H proton, is nearly 2 full  $pK$  units more acidic than the reported  $pK_a$  of 9.2 for the imino proton of native dG.<sup>11</sup>

**NMR Experiments.** The two-component strands of the sequence in Figure 1B were mixed in a 1:1 molar ratio by following the imino protons, and subsequently, 1D <sup>1</sup>H, <sup>1</sup>H/<sup>1</sup>H, <sup>1</sup>H/<sup>13</sup>C, and 2D <sup>1</sup>H/<sup>31</sup>P NMR spectra were acquired on the duplex in a buffer containing 20 mM sodium phosphate. A portion of the 2D <sup>1</sup>H/<sup>13</sup>C HSQC spectrum is shown in Figure S1A in the Supporting Information. The dispersion of the resonances was good, indicating an ordered three-dimensional structure. All nonexchangeable proton resonances were assigned to specific protons in the DNA molecule, except for a few of the H5'/H5'' protons, with a combination of 2D NMR techniques including <sup>1</sup>H/<sup>1</sup>H DQF-COSY, TOCSY, and NOESY, <sup>1</sup>H/<sup>13</sup>C HSQC and HMBC, and <sup>1</sup>H/<sup>31</sup>P heteronuclear COSY. The assigned proton and <sup>13</sup>C resonances are tabulated in the Supporting Information.

The 1D <sup>1</sup>H NMR spectrum acquired in H<sub>2</sub>O at pH 7.0, showing the downfield-shifted imino protons of guanine,

(5) (a) Liu, H.; Gao, J.; Lynch, S.; Maynard, L.; Saito, D.; Kool, E. T. *Science* **2003**, *302*, 868–871. (b) Hikishima, S.; Minakawa, N.; Kuramoto, K.; Fujisawa, Y.; Ogawa, M.; Matsuda, A. *Angew. Chem., Int. Ed.* **2005**, *44*, 596–598.

(6) (a) Gao, J.; Liu, H.; Kool, E. T. *J. Am. Chem. Soc.* **2004**, *126*, 11826–11831. (b) Liu, H.; Gao, J.; Kool, E. T. *J. Am. Chem. Soc.* **2005**, *127*, 1396–1402.

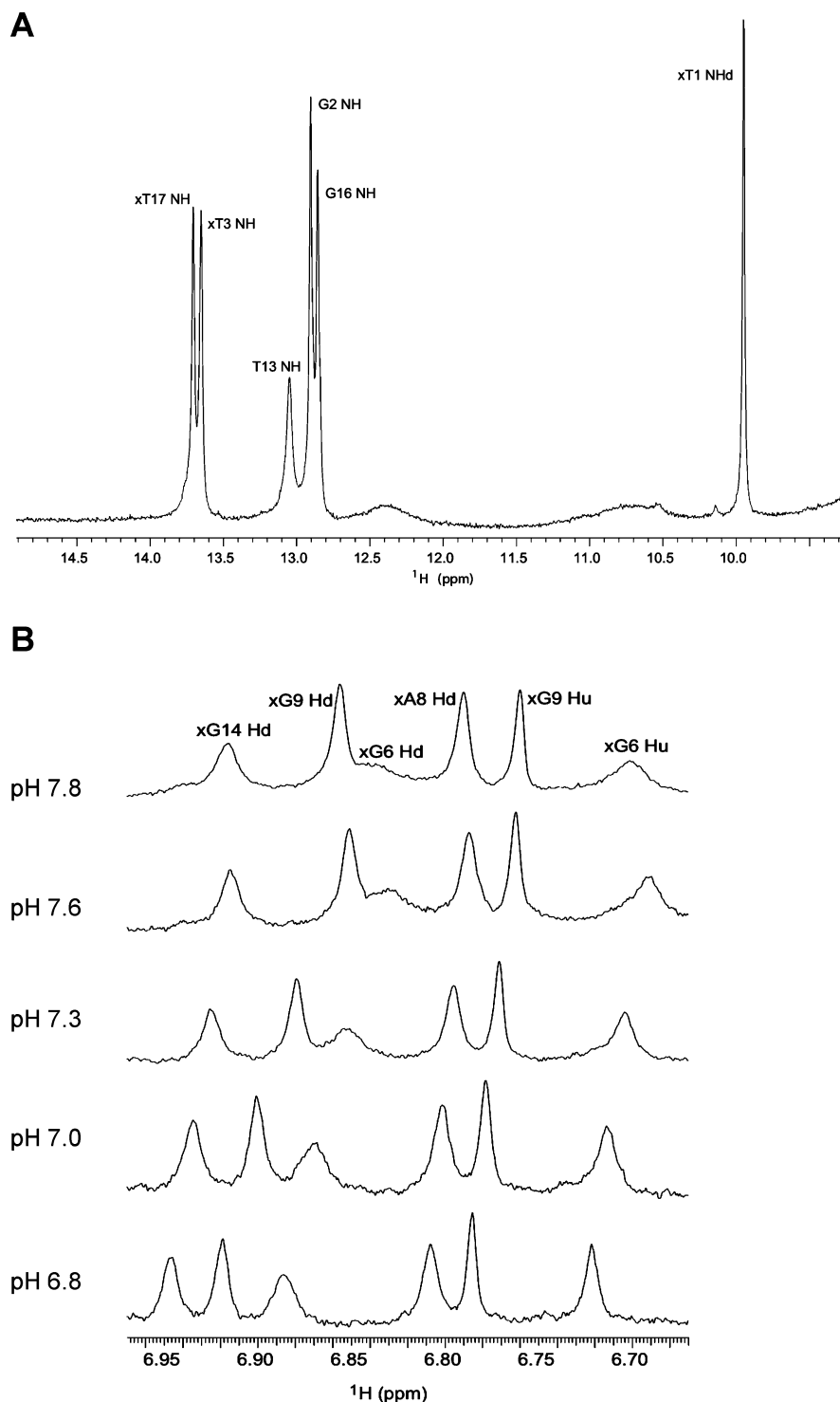
(7) (a) Lu, H.; He, K.; Kool, E. T. *Angew. Chem., Int. Ed.* **2004**, *43*, 5834–5836. (b) Lee, A. H. F.; Kool, E. T. *J. Am. Chem. Soc.* **2005**, *127*, 3332–3338.

(8) Gao, J.; Liu, H.; Kool, E. T. *Angew. Chem., Int. Ed.* **2005**, *44*, 3118–3122.

(9) Liu, H.; Lynch, S. R.; Kool, E. T. *J. Am. Chem. Soc.* **2004**, *126*, 6900–6905.

(10) Patterson, G. S. *J. Chem. Educ.* **1999**, *76*, 395–398.

(11) Chiang, Y.; Kresge, A. J. *Org. Biomol. Chem.* **2004**, *2*, 1090–1092.

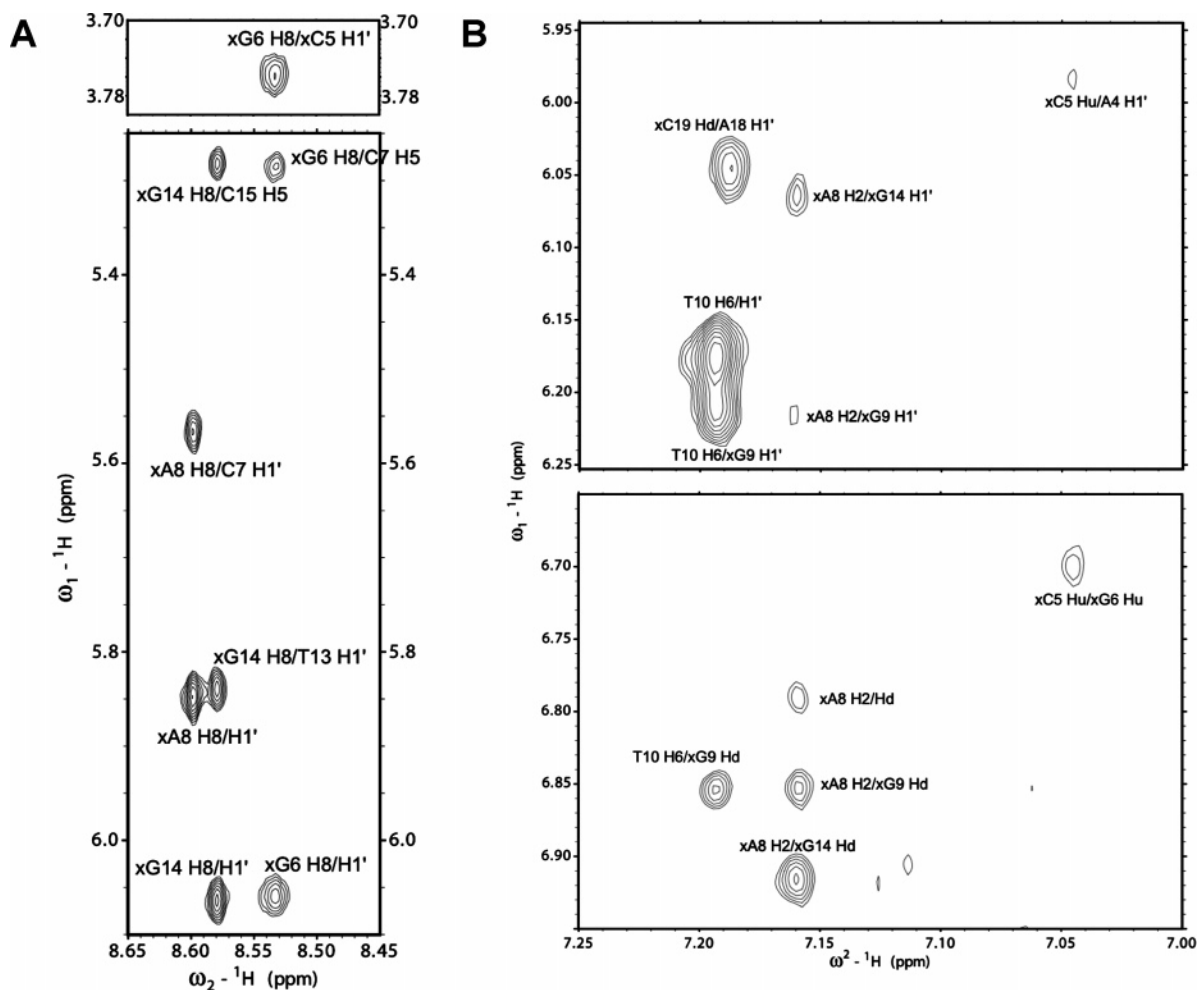


**Figure 2.**  $^1\text{H}$  1D spectra of the 10mer xDNA duplex in water: (A) spectrum acquired in 20 mM sodium phosphate buffer in  $\text{H}_2\text{O}$ , pH 7.0, showing the downfield-shifted imino proton resonances of G, T, and xT, with xG resonances missing due to exchange; (B) spectrum acquired in 20 mM sodium phosphate buffer in  $\text{D}_2\text{O}$  at various pH values, showing the pH dependence of the aromatic proton resonances of xG and those of protons close in space to xG.

thymine, and extended variants, is shown in Figure 2. Five imino resonances were observed to be sharp and were assigned in combination with  $^1\text{H}/^1\text{H}$  NOESY acquired in  $\text{H}_2\text{O}$ . A sixth sharp resonance in the downfield region of the NMR spectrum was assigned to a second NH resonance present on the x-pyrimidines; the other four such resonances of the x-pyrimidines were observed further upfield and are not shown.

Each Watson–Crick-like base pair of xDNA is expected to have an imino proton hydrogen-bonded to the complementary

base (Figure 1A); thus, one would expect to observe as many as 10 imino proton resonances. However, only five imino proton resonances were observed as sharp and assignable; the other five expected resonances were not assigned. The imino proton resonances for the two terminal base pairs were not observed, presumably due to rapid exchange with water; this result is normal for NMR spectra of DNA duplexes. The other three imino proton resonances that were not observed correspond to those of the xG bases. Broad resonances potentially belonging



**Figure 3.** Selected portions of the  $^1\text{H}/^1\text{H}$  NOESY spectrum of the decamer duplex acquired at 25 °C in 20 mM sodium phosphate buffer in  $\text{D}_2\text{O}$ , pH 7.0: (A) NOEs between the downfield aromatic protons and the  $\text{H}1'$  of the sugar or the  $\text{H}5$  of the pyrimidine bases, indicating stacking of nucleotides; (B) NOEs between the adenosine  $\text{H}2$  proton to aromatic or  $\text{H}1'$  protons on the opposite strand, indicating an intact helix.

**Table 1.** Structure Statistics and Atomic Rms Deviations

no. of distance constraints	
total	635
intranucleotide	356
internucleotide	279
interstrand	29
no. of dihedral constraints	80
no. of hydrogen bonds (WC base pairs)	25
final forcing energies for distance and dihedral constraints (kcal/mol)	$11.3 \pm 3.7$
av no. of NOE violations ( $>0.2 \text{ \AA}$ )	$<1$
maximum NOE violation ( $\text{\AA}$ )	0.25
av no. of dihedral violations ( $>2^\circ$ )	6
maximum dihedral violation (deg)	10
rmsd from distance constraints ( $\text{\AA}$ )	0.01
rmsd from dihedral constraints (deg)	0.67
deviations from idealized geometry	
bonds ( $\text{\AA}$ )	0.0050
angles (deg)	0.94
impropers (deg)	0.18
heavy-atom rmsd	
all (2–10, 11–19)	$1.18 \pm 0.11$
core residues (3–8, 13–18)	$0.80 \pm 0.11$

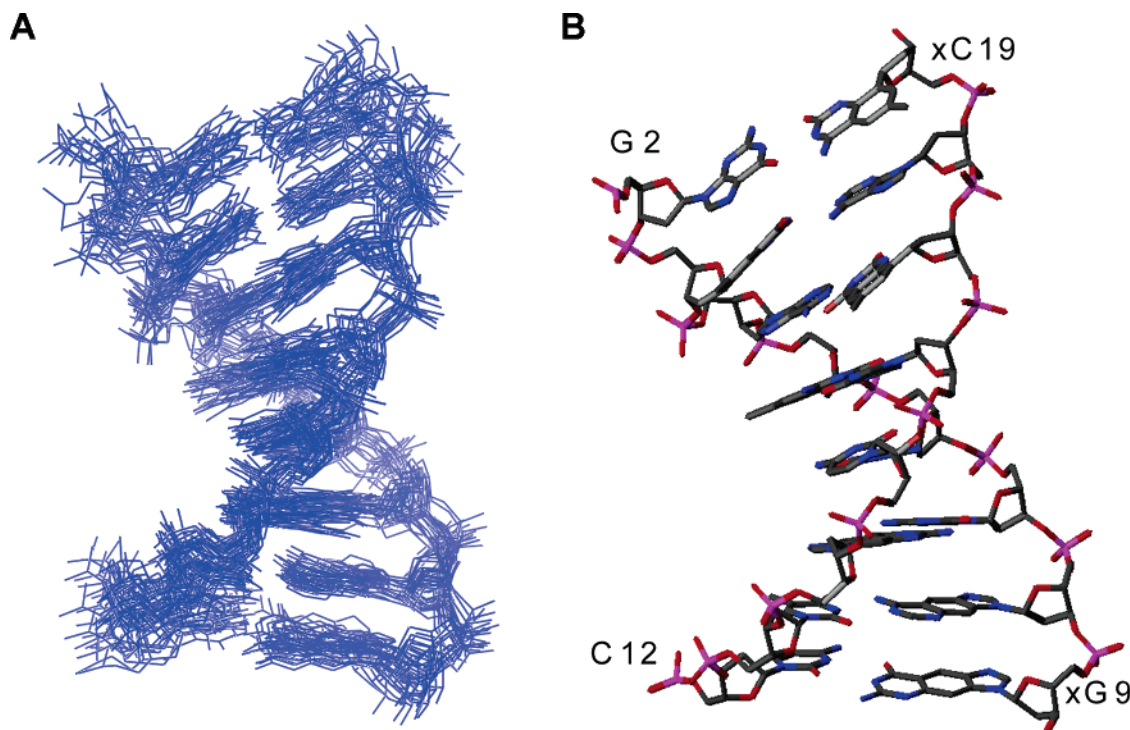
to the imino protons of xG were observed between 10.5 and 12.5 ppm (Figure 2).

Studies of free dxG in solution (above) showed a significantly lower  $\text{p}K_a$  than that of G. Assuming that the NH protons of xG's in the duplex were to have a lower  $\text{p}K_a$  as well, one would

expect broadening of the NH resonance due to exchange. In preliminary experiments, lowering the pH of the buffer below 7.0 caused some sharpening of the broad resonances; however, the other resonances broadened, and no additional assignments could be made.

To explore this absence of resonances further, 1D NMR and 2D NMR data were acquired in  $\text{D}_2\text{O}$  buffer at varied pH values (Figure 2B). The data showed that although many of the protons were unaffected by pH (in portions of the aromatic region not shown), resonances of xG and adjacent bases were clearly affected by pH in the 6.8–7.8 range. The change in chemical shift and broadening of the set of resonances were consistent with local conformational or structural change correlated to pH. As all of the observed changes were in either the proton resonances of the xG nucleotides or the resonances of protons on nucleotides close in space to an xG residue, this is consistent with protonation/deprotonation leading to the observed pH-induced broadening and change in chemical shift. This event would likely be from the N1 position of xG on the Watson–Crick face. The actual  $\text{p}K_a$  values of the three xG residues were not measurable with NMR, as pH values of 8.0 and above caused general broadening of the NMR spectrum (data not shown), making interpretation difficult. At the  $\text{p}K_a$ , the line width of the resonance should be the broadest, which implies





**Figure 4.** Structures of the xDNA helix containing 8 different bases: (A) superposition of 20 final structures after restrained molecular dynamics and energy minimization; (B) average structure of the xDNA duplex.

that the  $pK_a$  is 7.8 or higher for the imino proton for all three xG bases in the duplex.

Although the imino proton resonances were not observable for the xG bases, the helix is intact as confirmed by NOESY spectra acquired in  $D_2O$  buffer (Figure 3 and Figure S2 in the Supporting Information). NOEs were observed between the two strands of xDNA from xA8 on one strand and xG14 on the other strand. The H2 proton on xA is on the Watson–Crick-analogous face of the nucleotide and thus is normally close to the other strand in a Watson–Crick-like base pair. In addition to the cross-strand NOEs between xA8 and xG14, the expected NOEs indicating base stacking from successive nucleotides were also observed, including those between xC5 and xG6, xA8 and xG9, xG9 and T10, T13 and xG14, and xG14 and C15. Thus, the data confirm that although the xG hydrogen-bonded resonances are not observable, the three xG residues and their partners are located in the helix and are stacked as expected with the neighboring bases.

#### Structure of the Duplex Containing Eight Different Bases.

Structures of the decamer duplex were calculated from 635 NOEs, 80 experimentally determined dihedral constraints, and 25 hydrogen bonds for the Watson–Crick-like base pairs. The structure statistics for the set of 20 converged structures is presented in Table 1. The superposition of these structures is presented in Figure 4. The overall heavy-atom rmsd for the 10-nucleotide duplex is 1.18 Å. The primary reasons for disorder in the duplex are slight changes of helical bending, which are difficult to define by NMR due to the lack of long-range constraints, disorder at the terminal base pairs, which is common as there are fewer internucleotide NOEs to terminal nuclei, and the disorder correlated to the xG base. The resonances for the xG nucleotide as well as those of protons close in space to the xG imino proton were broad as shown in Figure 2B. Fewer

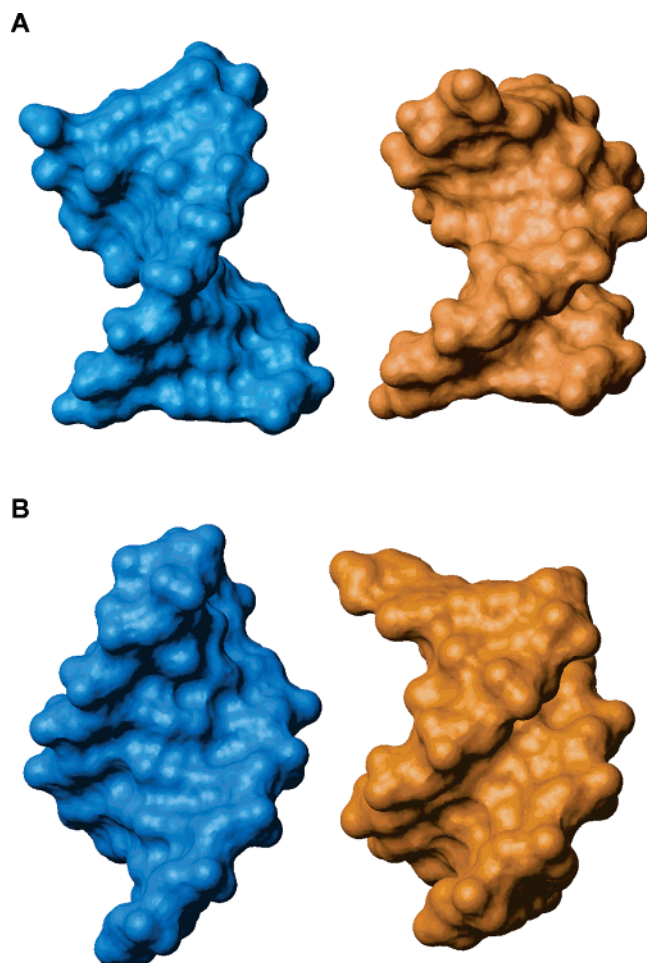
**Table 2.** Average Structural Parameters for xDNA and B-DNA

	xDNA	B-DNA
helix handedness	right	right
no. of bp's per repeating unit	1	1
no. of bp's per turn	12	10
helix twist (deg)	30	36
rise per bp (deg)	4.0	3.4
helix pitch (deg)	46	34
base pair inclination (deg)	24	2.4
diameter (Å)	20.5	18.4
X displacement from bp to helix axis (Å)	−1.8	−0.2
glycosidic bond orientation	<i>anti</i>	<i>anti</i>
sugar conformation	<i>2-endo</i>	<i>2-endo</i>
major groove depth (Å)	2.5	4.1
major groove width (Å)	12.4	11.4
minor groove depth (Å)	2.6	5.5
minor groove width (Å)	13.1	5.9

NOEs to these resonances were observed due to the putative exchange event, leading to fewer distance constraints near these residues.

The average structure of the xDNA duplex is shown in Figure 4B. The structure is a right-handed helix that is globally similar to that of B-DNA with the 20 bases in the *anti* conformation and the sugars in the *C2'-endo* conformation. Table 2 compares structural parameters of this xDNA structure and generic B-DNA. The most obvious differences are the width of the helix and the width and depth of the two grooves, particularly the minor groove. This marked difference in groove dimensions is obvious in surface models (Figure 5). In addition, there is a significant difference in the number of residues per turn (12 for xDNA versus 10 for B-DNA), which is correlated to the difference in the helix twist ( $36^\circ$  versus  $30^\circ$ ).

The structures of the four different base pair combinations from the average structure are shown in Figure 6. The Watson–Crick-analogous geometries are observed for each of the pairs. The most notable feature is the extensive stacking overlap of



**Figure 5.** Space-filling surface models showing the helix shape and grooves of xDNA (cyan, left) compared with B-DNA (orange, right): (A) view of the backbone with the major groove at the top; (B) minor groove view showing the exceptionally wide and shallow minor groove of xDNA.

xA and xG with each other. The x-purines within the same strand have extensive amounts of shared surface contact (see Figure 6B), such as in the case of xA8 and xG9. Even when situated in opposite strands, x-purines overlap to a significant degree when in adjacent pairs; examples are found with xA8 and xG14 and xG14 and xG6.

## Discussion

We previously reported the structure of a simple expanded DNA helix consisting of only two bases (xA and T).<sup>9</sup> Notable features of that structure included a well-defined xA–T hydrogen-bonded pair and a widened groove structure. However, the simple nature of the sequence of that structure left open the possibility that more complex sequences containing xDNA building blocks might form different structures. Indeed, in natural DNA it is known that while most sequences of DNA usually assemble into the B-type canonical helix, a number of simple sequences form very different structures. Examples include the parallel-stranded structures formed by A,T-DNA,<sup>12</sup> triple helices formed in purine stretches,<sup>13</sup> curved structures found with runs of adenines,<sup>14</sup> and quadruplexes formed in

G-rich DNA.<sup>15</sup> Interestingly, early evidence suggests that certain homo-x-pyrimidine sequences of xDNA may in fact form triple helices with homopurine DNAs.<sup>6b</sup>

The current study addresses the structure of the complete eight-base xDNA genetic set for the first time. The full sequence complexity of xDNA is approached in the current sequence by presenting four different base pairs (composed of eight different bases) in nine nearest neighbor contexts. Overall, the results confirm that the four xDNA pairs are, on average, hydrogen-bonded in the expected way, and that the xDNA helix is right-handed with pairs oriented roughly perpendicular to the helix, analogous to the previous simpler xDNA structure.<sup>9</sup> Compared with that earlier structure, the helix twist per pair and the helix diameter are almost the same (Table 2). However, there are some notable differences as well. There is a greater rise per base pair (4.0 Å vs 3.1 Å earlier) and greater helix pitch (46 Å vs 37 Å). The base pair inclination (24° in the current structure vs –0.5° before) is considerably greater. Finally, the grooves are shallower in the more complex sequence structure (2.5–2.6 Å vs 7–8 Å previously). This latter difference, and perhaps some of the other deviations as well, likely arises from the presence of xC–G and xT–A pairs, which are geometrically somewhat different from the xA–T pair that made up the earlier simple sequence (see Figures 1 and 7). The x-pyrimidines possess major groove methyl groups that make that groove shallower on average, and the x-pyrimidine geometry causes them to project further into the minor groove than do x-purines, effectively raising the floor of that groove as well.

Another substantial difference between the present eight-base xDNA structure and that of the earlier simple two-base system is the current observation of greater dynamics in the base pairs. We suggest two explanations for this. First is the ionization equilibrium and solvent exchange of the xG imino protons. To the extent that this base is ionized in the helix, this might promote pair opening and interaction with the solvent. Our observation of broad or missing NH protons of xG residues and pH-dependent line broadening near these bases is consistent with such motions. A second possible explanation is the small but significant difference in the geometries of pairs involving x-purines as compared with x-pyrimidines (Figure 7). This difference is likely to require small backbone and pair geometry adjustments when such pairs are adjacent, so some small degree of conformational shifts resulting from this would not be surprising. Either or both of these factors may contribute to the observed heavy-atom rmsd of 1.18 Å (Table 1), which is somewhat higher than that of the earlier (and simpler) two-base structure (0.84 Å).

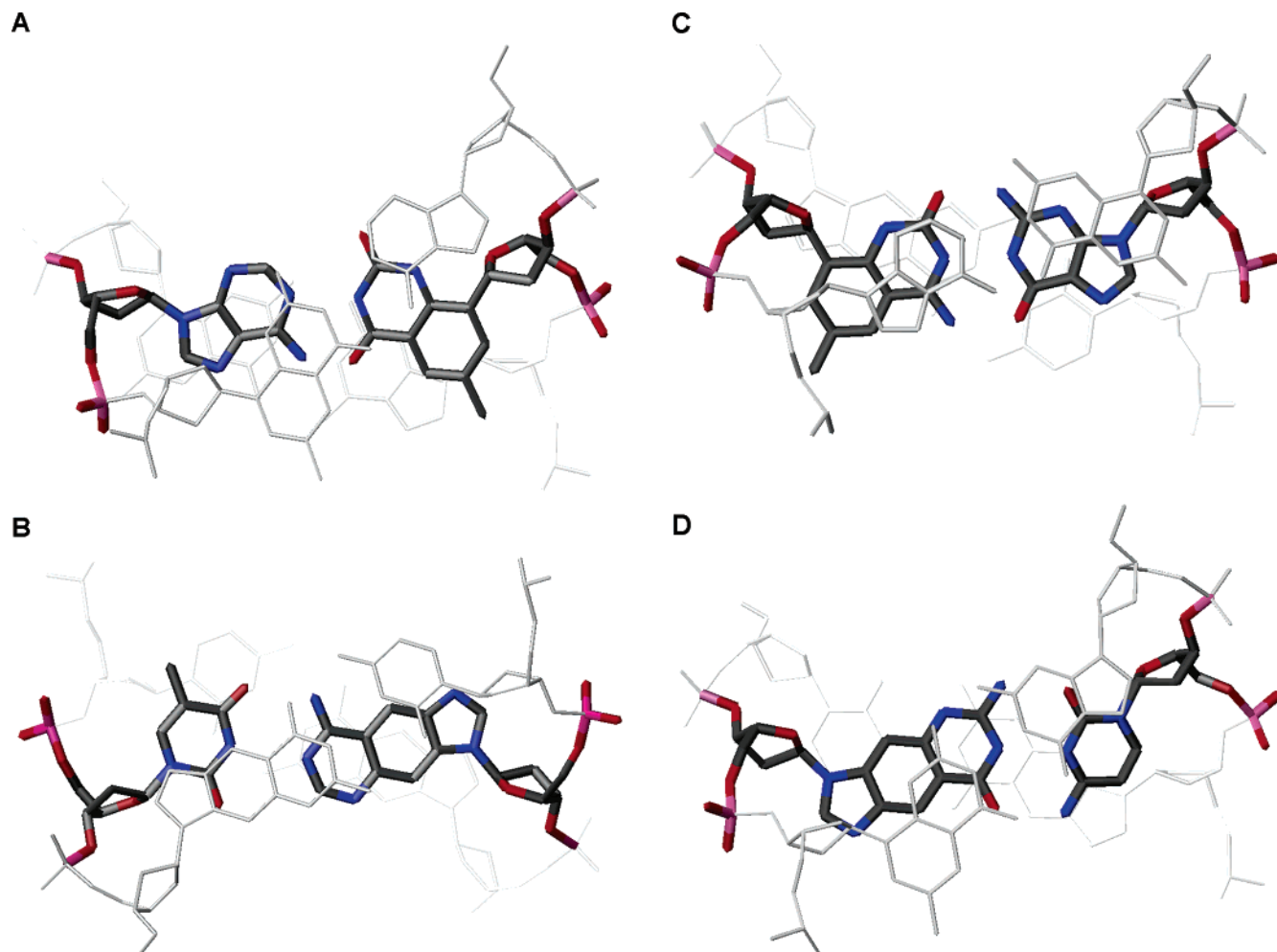
Overall, the current results confirm that mixed-sequence xDNA can pair and assemble sequence-specifically into a regular double helix in a predictable way. This specific self-assembly is an important feature of functioning genetic systems, which must store and transfer genetic information as a sequence of bases. DNA strands must be able to spontaneously form stable complexes after being denatured globally (such as in PCR) or locally (as during transcription). Moreover, the functioning of the natural genetic system in information transfer (such as during replication) requires a regular geometry of base pairs so that enzymatic copying can occur in a sequence-generic way. The

(12) Rippe, K.; Jovin, T. M. *Methods Enzymol.* **1992**, *211*, 199–220.

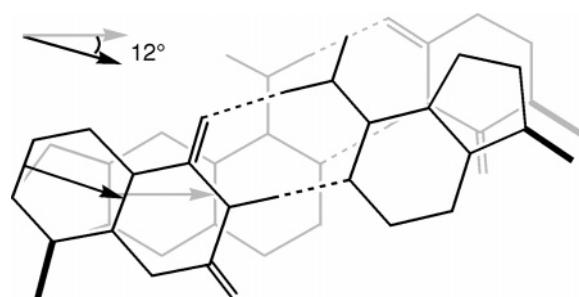
(13) (a) Morgan, A. R.; Wells, R. D. *J. Mol. Biol.* **1968**, *37*, 63–80. (b) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645–650.

(14) Koo, H.-S.; Wu, H.-M.; Crothers, D. M. *Nature (London)* **1986**, *320*, 501–506.

(15) Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. *Cell* **1989**, *59*, 871–880.



**Figure 6.** Structures of four different base pairs in the xDNA helix, with the adjacent nucleotides (gray) to illustrate base stacking arrangements: (A) A4-xT17; (B) T13-xA8; (C) xC5-G16; (D) xG6-C15.



**Figure 7.** Schematic overlay of xT-A (black) and xA-T (gray) base pairs drawn with conserved hydrogen bond lengths and angles, showing small geometric differences between benzo expansion of pyrimidines and purines in xDNA pairs. Vectors of expansion are depicted with arrows. Since C1'-C1' distances are constrained by the xDNA backbone, adjustments in hydrogen bond lengths and orientations are likely needed to compensate for the vector differences when such pairs are adjacent.

present work shows that the eight-base xDNA system may have most if not all of these features, which suggests the future possibility of a functioning, replicable genetic system using xDNA as the genetic material. Of course, it remains to be seen whether xDNA can in fact be replicated; this is expected to be a significant challenge since replicative DNA polymerases can be highly sensitive to the size of base pairs.<sup>16</sup>

On the practical level, this study shows that xDNA pairs can assemble in the predicted way, using complementarity of size, shape, and hydrogen bonding to assemble an antiparallel, regular helix. We have previously shown that strands composed entirely of the four expanded DNA bases can recognize natural DNA and RNA,<sup>8</sup> and the current results indicate that such recognition is structurally predictable in the same way as natural DNA assembly. Because xDNA bases are inherently fluorescent and form highly stable helices, this could lead to future development of useful tools for probing and detection of nucleic acids.

## Experimental Section

**Oligonucleotide Synthesis and Purification.** The deoxynucleoside phosphoramidite derivatives of the four benzo-homologated nucleosides were prepared as described previously.<sup>17</sup> The two 10-mer strands of the present duplex were prepared as reported.<sup>8</sup> Briefly, sequences contained a 3'-phosphate group using a "3'-phosphate-ON" controlled pore glass support (Glen Research). Oligonucleotides were synthesized using an Applied Biosystems 392 synthesizer in trityl-off mode using standard  $\beta$ -cyanoethyl phosphoramidite chemistry. Two  $1 \mu\text{mol}$  syntheses of each sequence were performed and the products combined for each. Oligonucleotides were deprotected and removed from the solid support with concentrated ammonia. They were purified by reversed-phase HPLC and quantitated by UV absorption.

(16) Kim, T. W.; Delaney, J. C.; Essigmann, J. M.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15803-15808.

(17) (a) Liu, H.; Gao, J.; Saito, D.; Maynard, L.; Kool, E. T. *J. Am. Chem. Soc.* **2004**, *126*, 1102-1109. (b) Liu, H.; Gao, J.; Kool, E. T. *J. Org. Chem.* **2004**, *70*, 639-647.



**NMR Structural Studies.** NMR experiments were acquired on either a Varian Inova 600 MHz NMR instrument or a Varian Inova 500 MHz NMR instrument equipped with triple-resonance and  $z$ -gradient or triple-axis gradient capabilities. NMR spectra were acquired at 25 °C in 20 mM sodium phosphate, pH 7.0, exchanged into 99.996% D<sub>2</sub>O, or at 2, 15, and 25 °C in 90% H<sub>2</sub>O/10% D<sub>2</sub>O in 20 mM sodium phosphate buffer, pH 7.0. <sup>1</sup>H resonances were assigned with standard methods using a combination of DQF-COSY, TOCSY with mixing times of 20 and 80 ms, <sup>1</sup>H/<sup>13</sup>C HSQC, <sup>1</sup>H/<sup>31</sup>P heteronuclear COSY, <sup>1</sup>H/<sup>13</sup>C HMBC, NOESY, and SS-NOESY experiments. The spectral width in the <sup>1</sup>H dimension for all experiments in D<sub>2</sub>O was 10 ppm. <sup>1</sup>H/<sup>13</sup>C HSQC was acquired with 2048 points in the  $t_2$  dimension by 128 real points in the  $t_1$  dimension with a recycle delay of 2.5 s and a carbon spectral width of 70 ppm for the aromatic resonances and 2048 by 256 with a 1.8 s recycle delay and a spectral width of 40 ppm. DQF-COSY was acquired with 16 scans of 8192 points in the  $t_2$  dimension by 512 points in the  $t_1$  dimension with an acquisition time of 0.682 s and a recycle delay of 0.75 s. TOCSY experiments were acquired with 16 scans of 4096 points in  $t_2$  by 400 points in  $t_1$  with an acquisition time of 0.341 s and a recycle delay of 0.75 s. NOESY experiments in D<sub>2</sub>O were acquired with mixing times of 50, 100, 150, and 300 ms with 32 scans of 4096 points in  $t_2$  and 400 in  $t_1$  with an acquisition time of 0.341 s and a recycle delay of 1.5 s. SS-NOESY experiments were acquired in 90% H<sub>2</sub>O with mixing times of 75 and 250 ms, with 160 scans of 4096 by 256 points, an acquisition time of 0.179 s, and a recycle delay of 1.5 s with a 19 ppm wide spectral width.

NMR data were processed using Varian VNMR 6.1C software. Display of NMR data, coupling constant measurement, and calculation of NOE volumes were accomplished in the NMR display program using SPARKY 3.<sup>18</sup>

The protons on the deoxyribose residues were distinguished by type with <sup>1</sup>H/<sup>13</sup>C HSQC acquired in natural abundance and separated into nucleotide spin systems with TOCSY and assigned with DQF-COSY. The deoxyribose protons were correlated to the base by <sup>1</sup>H/<sup>13</sup>C HMBC when possible or with strong intranucleotide NOEs between H2' and H6 for thymidine and cytidine, H2' and H8 for adenosine and guanosine, and H1' and Hd and H2' and H8 for xA (see the Supporting Information for numbering). The protons on the aromatic ring of the eight different nucleotides could be distinguished from each other by the <sup>13</sup>C chemical shift as observed in <sup>1</sup>H/<sup>13</sup>C HSQC. These proton resonances were assigned by a combination of TOCSY correlation to each other through-bond or NOESY correlation through-space. Sequential assignments were made through many NOEs observed between sequential nucleotides including protons on the deoxyribose and the protons on the sequential base and protons on one base and protons on the sequential base. The imino protons of T, xT, G, and xG could be assigned on the basis of imino–imino NOEs between sequential base pairs and intense NOEs between the T and xT imino and the H2 proton of xA (see the numbering in the Supporting Information) or the H2 proton of A. Additional NOEs were observed between the H2 of xA or the H2 of A and imino protons of sequential base pairs to confirm the assignments.

**Structure Calculation.** Structures of the decamer DNA duplex were calculated on an SGI Octane workstation with restrained molecular dynamics followed by energy minimization with the program CNS.<sup>19</sup> The nonnatural bases were created first by building them with InsightII (Accelrys, San Diego, CA) and outputting coordinate and parameter files of bond lengths, angles, and improper angles into CNS. These bases were combined with the deoxyribose to make a coordinate file of the four nonnatural bases. The normal parameter file for deoxynucleotides in CNS was modified to include the nonnatural bases. The nonnatural nucleotides were then connected to the natural ones in the proper sequence of the two strands in the CNS program.

One hundred random starting structures were created by giving different initial velocities to an elongated single strand of each of the two complementary oligonucleotide sequences. These structures were subjected to a simulated annealing protocol with restrained torsion angle dynamics using a force field of bond lengths, bond angles, improper angles, repulsive van der Waals potentials, and experimental distance and torsion angle constraints. NOE force constants were set at 50 kcal/(mol Å<sup>2</sup>); torsion angle force constants were varied from 10 to 50 kcal/(mol Å<sup>2</sup>). The high-temperature annealing was done with 4000 cycles of molecular dynamics for 60 ps at 20000 K with low values of interatomic repulsion, followed by 1000 steps of torsion angle dynamics with cooling to 2000 K and slow cooling with 2000 steps of Cartesian dynamics to 300 K, increasing the dihedral force constant from 10 to 50 kcal/(mol Å<sup>2</sup>). An initial energy minimization with 200 steps with both the NOE and dihedral force constants at 50 kcal/(mol Å<sup>2</sup>) was accomplished. Of 100 initial structures, the 20 lowest energy structures were subjected to a second stage of 2500 cycles of molecular dynamics at 300 K; this was followed by a final energy minimization that included all of the above constraints in addition to attractive Lennard-Jones potentials without electrostatic potentials for 3000 steps. The final structures were superimposed and displayed with either the program MOLMOL<sup>20</sup> or the program Pymol.<sup>21</sup>

Distance restraints for nonexchangeable protons were assigned on the basis of analysis of the cross-peak intensity in D<sub>2</sub>O NOESY experiments with mixing times of 50, 100, 150, and 300 ms and SS-NOESY experiments in H<sub>2</sub>O with 75 and 250 ms mixing times. NOEs were assigned as strong (1.8–3.0 Å), medium (2.3–4.5 Å), and weak (3.0–6.0 Å). Hydrogen bonds for the Watson–Crick-like base pairs were assigned on the basis of intense NOEs observed between the thymine or x-thymidine imino proton and x-adenosine H2 or adenosine H2 proton or the guanosine or imino proton and the x-cytidine amino protons, as well as the downfield chemical shift of the imino proton resonance. The x-guanosine–cytosine base pairs and terminal base pairs were modeled as Watson–Crick-like pairs even though the imino protons could not be observed due to solvent exchange.

Dihedral constraints were assigned on the basis of analysis of DQF-COSY and <sup>1</sup>H/<sup>31</sup>P heteronuclear COSY. The sugar pucker was constrained on the basis of comparison of <sup>3</sup>J<sub>H1'H2'</sub>, <sup>3</sup>J<sub>H1'H2''</sub>, and <sup>3</sup>J<sub>H3'H4'</sub> values. All nucleotides were determined to be C2'-endo except for the terminal thymidine and x-thymidine, which were observed to be in conformational exchange. The backbone torsion angle  $\beta$  was estimated on the basis of measurement of <sup>3</sup>J<sub>H5'P</sub>, <sup>3</sup>J<sub>H5'P'</sub>, and <sup>4</sup>J<sub>H4'P</sub> and assigned a range of  $\pm 30^\circ$ . The backbone torsion angle  $\gamma$  was estimated on the basis of measurement of <sup>3</sup>J<sub>H4'H5'</sub>, <sup>3</sup>J<sub>H4'H5''</sub>, and <sup>4</sup>J<sub>H4'P</sub> and assigned a range of  $\pm 30^\circ$ .

**Helical Parameters.** The helical parameters and torsion angles for the decamer duplex were determined using the program CURVES 5.3 by inputting the structures in pdb format into the program after final energy minimization.<sup>22</sup>

**Acknowledgment.** This work was supported by the U.S. National Institutes of Health (Grant GM63587). Dedicated to the memory of Professor Nelson Leonard.

**Supporting Information Available:** Additional 2D spectral data and chemical shift tables. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA065606N

(18) Goddard, T. D.; Huang, C. C.; Ferrin, T. E. *Structure* **2005**, *13*, 473–482.  
(19) Brünger, A. T.; Adams, P. D.; Clore, G. M.; Delano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr., D: Biol. Crystallogr.* **1998**, *54*, 905–921.

(20) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51–55.  
(21) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2002; <http://pymol.sourceforge.net>.  
(22) Lavery, R.; Sklenář, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63–91.