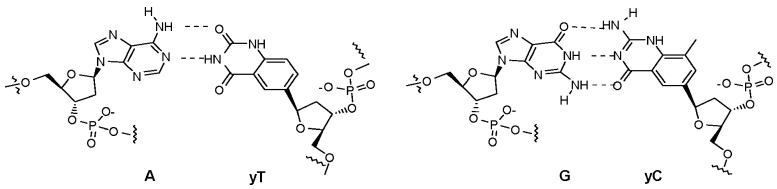
Article

A New Four-Base Genetic Helix, yDNA, Composed of Widened Benzopyrimidine–Purine Pairs

Alex H. F. Lee, and Eric T. Kool

J. Am. Chem. Soc., **2005**, 127 (10), 3332-3338• DOI: 10.1021/ja0430604 • Publication Date (Web): 18 February 2005



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 12 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 02/18/2005

A New Four-Base Genetic Helix, yDNA, Composed of Widened Benzopyrimidine−Purine Pairs

Alex H. F. Lee and Eric T. Kool*

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

Received November 17, 2004; E-mail: kool@stanford.edu

Abstract: We describe the properties of stable DNA-like self-assembled helices composed entirely of base pairs involving two new size-expanded pyrimidines. We term this new helix geometry "yDNA" (an abbreviation of "wide DNA"). The new pyrimidine analogues, yT and yC, are increased in size by benzo-homologation and have a geometry that is distinct from previous size-expanded pyrimidines. The yT and yC deoxyribosides were incorporated into oligodeoxynucleotides designed to form four pairs: yT−A, A−yT, yC−G, and G−yC. Helices were characterized by thermal denaturation, mixing data, and circular dichroism spectra. Results showed that highly stable double-stranded helices were formed in several sequence contexts. The data further showed that yT and yC could be segregated onto one strand and used to bind to natural strands of DNA with high sequence selectivity. The combination of yC, yT, G, and A make up a new selective, self-assembling four-base genetic pairing system that functions in many respects like natural DNA, but which is structurally distinct. The results establish that multiple variants of size-expanded DNA-like helices are feasible and suggest the possibility of a future eight-base genetic system based on the yDNA geometry. Finally, the high binding selectivity, affinity, and fluorescence of yDNA strands may yield useful applications in detection of nucleic acid sequences.

Introduction

We have recently undertaken a series of studies aimed at designing new genetic systems in which all base pairs are larger than those of natural DNA.[1−6] The design and study of such DNA-mimicking systems serves as a test of the current knowledge of the basic principles that enable the self-assembly and function of DNA itself. In addition, size-expanded bases and nucleotides may serve as useful tools in biophysical studies of DNA and proteins that interact with it. Finally, because of their exceptional hybridization and fluorescence properties, oligomeric strands of size-expanded bases might find useful applications in new biotechnologies as well.

For the concept of base pairs larger than the standard Watson−Crick geometry, a number of previous studies employing purine−purine pairs merit mention. Such pairs, if oriented *anti*−*anti*, can possibly take on a dimension wider than pyrimidine−purine Watson−Crick pairs do. For example, Benner et al. have reported studies characterizing purine−purine pairs in otherwise natural DNA duplexes,[7] although only single, isolated Pu−Pu pairs were examined. Studies with multiple purine−purine pairs have been pursued by Seela et al. and by Eschenmoser et al., who described helix formation by purine oligonucleotide strands paired with other purine strands.[8] In one example, a hexamer containing 5-aza-7-azaguanine was shown to form a stable duplex with a G-containing strand, resulting in a helix proposed to be parallel-stranded with base pairs oriented *anti*−*anti*.[8a] Finally, Matsuda et al. have recently reported non-natural pairs meant to form four hydrogen bonds, and which may adopt a geometry wider than natural pairs. Up to three consecutive pairs of this kind were incorporated into natural DNA.[9] Such bases and pairs have not been studied with complete replacement of all natural pairs, which would be needed to form a size-expanded helix.

To date, only a benzo-fusion design for expansion of nucleobases has led to fully expanded-size antiparallel duplex formation.[1,4,6] Our original approach to the design of size-expanded purines and pyrimidines to form larger helices was the linear extension of length by insertion of a benzene ring, yielding 2.4 Å of added size, in analogy to early studies of *lin*-benzo-ATP analogues by Leonard.[10] Our recent work yielded the expanded DNA (xDNA) nucleosides (dxT and dxA have been reported to date[1,2]) and helices containing them. In part because of their larger size, the xDNA nucleosides are fluo-

(1) Liu, H.; Gao, J.; Lynch, S. R.; Saito, Y. D.; Maynard, L.; Kool, E. T. *Science* **2003**, *302*, 868−871.
(2) Liu, H.; Gao, J.; Maynard, L.; Saito, Y. D.; Kool, E. T. *J. Am. Chem. Soc.* **2004**, *126*, 1102−1109.
(3) Liu, H.; Lynch, S. R.; Kool, E. T. *J. Am. Chem. Soc.* **2004**, *126*, 6900−6905.
(4) Gao, J.; Liu, H.; Kool, E. T. *J. Am. Chem. Soc.* **2004**, *126*, 11826−11831.
(5) Liu, H.; Gao, J.; Kool, E. T. *J. Org. Chem.* **2005**, *70*, 639−647.
(6) Lu, H.; He, K.; Kool, E. T. *Angew. Chem., Int. Ed.* **2004**, *43*, 5834−5836.
(7) Geyer, C. R.; Battersby, T. R.; Benner, S. A. *Structure* **2003**, *11*, 1485−1498.
(8) (a) Seela, F.; Melenewski, A.; Wei, C. *Bioorg. Med. Chem. Lett.* **1997**, *17*, 2173−2176. (b) Groebke, K.; Hunziker, J.; Fraser, W.; Peng, L.; Diederichsen, U.; Zimmermann, K.; Holzner, A.; Leumann, C.; Eschenmoser, A. *Helv. Chim. Acta* **1998**, *81*, 375−474. (c) Seela, F.; Wei, C.; Melenewski, A.; Feiling, E. *Nucleosides Nucleotides* **1998**, *17*, 2045−2052.
(9) Minakawa, N.; Kojima, N.; Hikishima, S.; Sasaki, T.; Kiyosue, A.; Atsumi, N.; Ueno, Y.; Matsuda, A. *J. Am. Chem. Soc.* **2003**, *125*, 9970−9982.
(10) Leonard, N. J. *Acc. Chem. Res.* **1982**, *15*, 128−135.

10.1021/ja0430604 CCC: $30.25

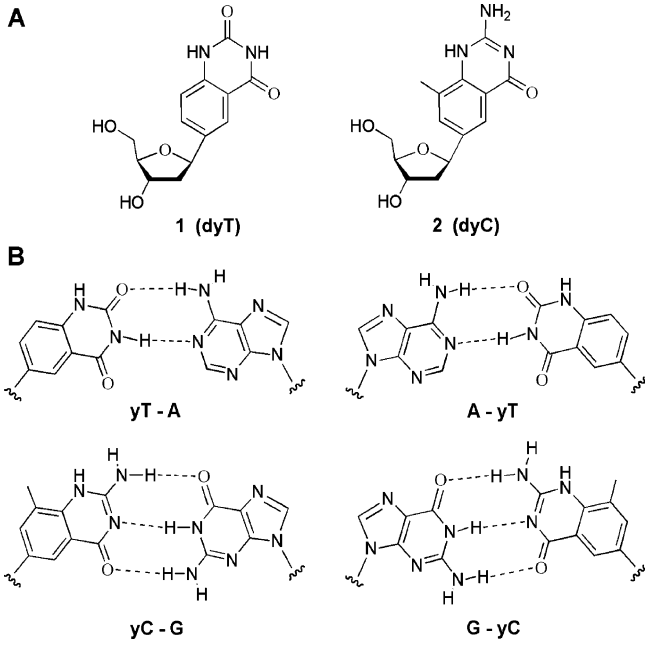


Figure 1. Structures of widened nucleoside analogues (A) and proposed structures of widened base pairs (B) in this study.

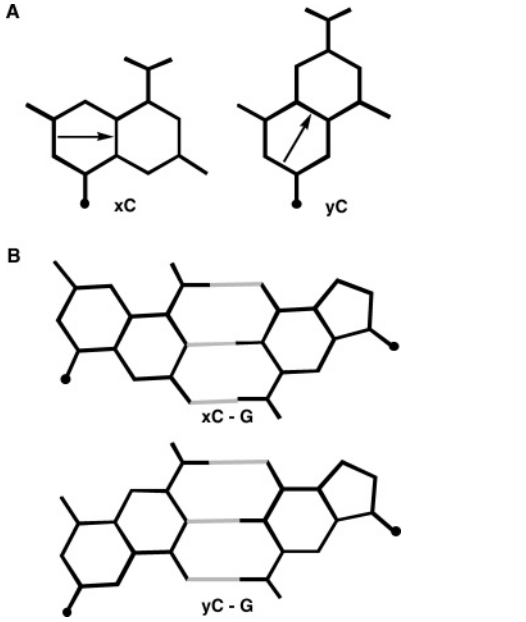


Figure 2. Designs of widened (yDNA) nucleobases and base pairs, with comparison to an earlier expanded (xDNA) geometry.[1] Examples shown involve homologues of cytosine; thymine homologues are exactly analogous. (A) Illustration of vectors of extension (denoted by arrows) conferred by benzo-homologation. The ball at the end of the vertical bond represents the glycosidic carbon of deoxyribose. (B) Illustration of the designed base pair structure for yDNA as compared to the earlier xDNA design. Glycoside bonds were kept in the same orientation for both cases.

rescent, and they stack with high affinity on neighboring DNA sequences.[3] Single substitutions of xDNA base pairs in natural DNA were shown to be somewhat destabilizing, because of the mismatch in size with the neighboring DNA. However, when strands were designed to form helices in which *every* base pair was expanded, the resulting duplexes were exceptionally thermally stable.[1,5]

Studies of the properties of that initial size-expanded helix design are ongoing. However, examination of models suggested to us that other designs might be equally viable for extension of base pairs. An early study introduced a new benzo-homologated design for pyrimidines,[11] termed "yDNA" (for "wide DNA"). The geometry of these widened pyrimidines (yT and yC, Figure 1) is different from previous extended pyrimidines (denoted xT and xC) by their vector of extension (Figure 2). The length of extension of the bases in yT and yC is essentially the same as that in xT and xC, at 2.4 Å, but the extension vector orientation is altered by 60°. Simple models of the expected base pairs (Figure 2) show that the overall effects of these two extensions on base pair length are similar, but are expected to yield different groove depths. More specifically, with the yDNA pairs, a putative helix might be expected to have a shallower major groove and deeper minor groove than the xDNA helix does.[5]

The syntheses of the yT and yC nucleosides were described recently.[11] When substituted at single positions in the context of natural DNA, the yT and yC bases were shown to be energetically neutral to weakly destabilizing, most likely because base pairs involving them are mismatched in size next to natural pairs, leading to a distortion of the helical backbone. Nevertheless, they were shown to stack strongly relative to their natural counterparts.

Here we conduct the first test of the ability of these new benzopyrimidines to assemble a designed four-base genetic system composed only of widened base pairs. The results show that such complexes do in fact form with the new pyrimidine analogues and that they can be considerably more stable and sequence-selective than natural DNA is. Possible applications of yDNA strands are highlighted, and the implications of the results to design artificial genetic systems are discussed.

Results

Pairing Stability. Previous results showed that the new yC and yT nucleobase analogues (Figure 1) were destabilizing when substituted in natural DNA, presumably because of the mismatch in size arising from these large bases relative to the surrounding natural DNA.[11] In the current study we wished to test whether they could form helical DNA-like complexes if *all* base pairs could adopt the widened geometry. This question was tested in a number of contexts, both in self-complementary sequences and in non-self-complementary oligomers designed to form antiparallel duplexes (Table 1). We tested sequences containing only yT−A and A−yT pairs and sequences containing only yC−G and G−yC pairs. Finally, we examined a number of sequences that were designed to contain all four of these pairs together. For comparison we made measurements with natural DNA helices having the analogous sequence (i.e., T replacing yT and C replacing yC). Note that, technically, yT is a widened analogue of U, and yC is an analogue of 5-methyl-C (5mC), so we also tested, for the sake of completeness, control sequences in which U replaces yT and 5mC replaces C. Helical stability for these yDNAs and controls was measured by thermal denaturation experiments monitored by absorbance changes in the UV spectrum.

Initial experiments were aimed at testing each of the widened bases alone in fully widened helices. Table 1 shows a number of sequences that were designed to form yDNA helices involving only yC−G and G−yC pairs, and Figure 3 gives examples of

(11) Lee, A. H. F.; Kool, E. T. *J. Org. Chem.* **2005**, *70*, 132−140.

Table 1. Thermal Melting Data and Estimated Free Energies for yDNA Duplexes[a]

sequence	T_m[b] (°C)	$\Delta G°_{37}$[c] (kcal/mol)
5'- ^{y}C ^{y}C ^{y}C G ^{y}C G G G$_p$ 3'-$_p$G G G ^{y}C G ^{y}C ^{y}C ^{y}C	71.1 ± 0.5	-14.0 ± 1.5
5'- C C C G C G G G$_p$ 3'-$_p$G G G C G C C C	53.4 ± 0.5	-10.1 ± 0.3
5'- ^{m}C ^{m}C ^{m}C G ^{m}C G G G$_p$ 3'-$_p$G G G ^{m}C G ^{m}C ^{m}C ^{m}C	57.8 ± 0.5	-10.3 ± 0.8
5'- ^{y}C ^{y}C ^{y}C G G ^{y}C G$_p$ 3'-$_p$G G G ^{y}C ^{y}C G ^{y}C	64.7 ± 0.5	-13.3 ± 0.6
5'- C C C G G C G$_p$ 3'-$_p$G G G C C G C	44.2 ± 0.5	-9.8 ± 0.2
5'- ^{m}C ^{m}C ^{m}C G G ^{m}C G$_p$ 3'-$_p$G G G ^{m}C ^{m}C G ^{m}C	52.3 ± 0.5	-11.4 ± 0.2
5'- ^{y}C G ^{y}C G ^{y}C G$_p$ 3'-$_p$G ^{y}C G ^{y}C G ^{y}C	55.7 ± 0.5	-9.1 ± 0.3
5'- C G C G C G$_p$ 3'-$_p$G C G C G C	44.4 ± 0.5	-8.9 ± 0.5
5'- ^{m}C G ^{m}C G ^{m}C G$_p$ 3'-$_p$G ^{m}C G ^{m}C G ^{m}C	47.3 ± 0.5	-8.8 ± 0.5
5'- ^{y}T ^{y}T ^{y}T A A ^{y}T A ^{y}T A A$_p$ 3'-$_p$A A A ^{y}T ^{y}T A ^{y}T A ^{y}T ^{y}T	58.2 ± 0.5	-15.6 ± 2.1
5'- T T T A A T A T A A$_p$ 3'-$_p$A A A T T A T A T T	15.2 ± 0.5	-4.3 ± 0.7
5'- U U U A A U A U A A$_p$ 3'-$_p$A A A U U A U A U U	< 10	--
5'- ^{y}C ^{y}C G ^{y}T ^{y}T A A ^{y}C G G$_p$ 3'-$_p$G G ^{y}C A A ^{y}T ^{y}T G ^{y}C ^{y}C	63.5 ± 0.5	-12.4 ± 0.8
5'- C C G T T A A C G G$_p$ 3'-$_p$G G C A A T T G C C	24.1 ± 0.5	-6.6 ± 0.2
5'- ^{m}C ^{m}C G U U A A ^{m}C G G$_p$ 3'-$_p$G G ^{m}C A A U U G ^{m}C ^{m}C	38.2 ± 0.5	-7.6 ± 0.1
5'- ^{y}C ^{y}C ^{y}T G A ^{y}C G$_p$ 3'-$_p$G G A ^{y}C ^{y}T G ^{y}C	44.6 ± 0.5	-9.4 ± 0.1
5'- C C T G A C G$_p$ 3'-$_p$G G A C T G C	27.1 ± 0.5	-7.0 ± 0.1
5'- ^{m}C ^{m}C U G A ^{m}C G$_p$ 3'-$_p$G G A ^{m}C U G ^{m}C	32.1 ± 0.5	-7.7 ± 0.1

[a] Native DNA controls are shown for each sequence. Conditions: 100 mM NaCl, 10 mM $MgCl_2$, and 10 mM PIPES·Na (pH 7.0). [b] T_m values are for 5.0 μM oligonucleotide. [c] Averages of values from van't Hoff and curve fitting methods.

thermal denaturation curves for these same sequences. The data show that all three sequence contexts involving only yC (two are self-complementary, while one is not) yielded well-behaved, apparently two-state complexes. All three were considerably more thermally stable than the DNA controls, with melting temperature (T_m) values 8.4−17.7 °C higher. We also measured curve fits and used van't Hoff plots at varied concentration to estimate free energy values for the helices (Table 1). Note that these estimates should be considered preliminary, since the van't Hoff method relies on the assumption that ΔCp is small, an assumption that is not yet tested (see Discussion). Given this caution, the values for the first two sequences showed greater stability than the controls by ca. 2−4 kcal/mol. The third case, a hexamer, gave estimated free energy values that fall within experimental error of one another. The yT nucleoside was also tested; this fourth sequence is shown in Table 1. In this decamer context, the yDNA sequence showed dramatic stabilization compared to the DNA controls, with a T_m 43 °C higher and an estimated free energy that was 11 kcal/mol more favorable.

We also tested two different sequence contexts that contain both yC and yT in mixed combinations with the partners G and A, respectively (Table 1 and Figure 3). The data showed a high degree of stabilization of the yDNA complexes compared to the controls, with T_m values as much as 39 °C higher and estimated free energies that were 2−6 kcal/mol more favorable.

Pairing Selectivity. We then tested whether the yT and yC nucleosides would display selectivity when paired opposite mismatched bases. This was examined by preparing an all-yDNA strand and hybridizing it with all-DNA strands containing a fully complementary sequence or with single mismatches opposite yT or yC at centrally located positions. The data are shown in Table 2, which gives the experimental stabilities for the fully matched case and for the six mismatched cases (entries 1−7). Also given in the table are the controls for the same experiments (entries 10−16), using an all-DNA strand of the same sequence to target the same series of matched and mismatched octamer oligonucleotides.

The results showed that, in this short sequence context, the natural DNA control strand gave relatively high selectivity against mismatches. For example, for mismatches opposite T (Table 2, entries 11−13) there was a resulting drop in T_m of 9−17 °C. This selectivity is compared graphically in Figure 4. Mismatches against C were also apparently strongly destabilizing: all three mismatches showed no cooperative melting transition, suggesting a T_m value of 10 °C or lower (Table 2).

After these measurements with the control, the yDNA strand was then tested for its sequence selectivity in the same context. The data showed (Table 2) that selectivity was higher than that of the previous DNA control. For the mismatches opposite yT, the T_m values dropped by 15−27 °C. Figure 4 shows a histogram of the T_m differences of these mismatched sequences as compared to those of the natural DNA control. The ordering of mismatch stability is the same in both cases (A ≫ G > T ≥ C). This suggests the possibility that yT−G forms a wobble geometry analogous to T−G. The selectivity of yC was also apparently high; in this case (as for the control DNA oligomer) the T_m values fell too low to be observed for the single mismatches (<10 °C). Since the T_m value of the yDNA fully matched sequence was higher (by 15 °C) than that of the DNA control, we surmise that the yDNA selectivity is higher than that of the natural DNA control.

Preferences for Strand Direction and DNA vs RNA Backbone. yDNA double helices necessarily must adopt some conformational differences as compared to DNA, to adjust to the larger glycosidic distance. Thus, one cannot automatically assume a preference for antiparallel strand orientation as seen by natural DNA. Indeed, some sequences of natural DNA can form parallel-oriented helices, albeit with lower stability than the antiparallel ones.[12] Thus, we carried out an experiment in the mixed yT,yC sequence context to evaluate strand orientation preference for yDNA. The data are given in Table 2, entries 8 and 17, with comparison to the antiparallel versions (entries 1 and 10). The results showed that the parallel orientation for both the yDNA strand and the control DNA strands did not show a transition above 10 °C in this sequence context. This suggests that there is a strong preference for antiparallel orientation for

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

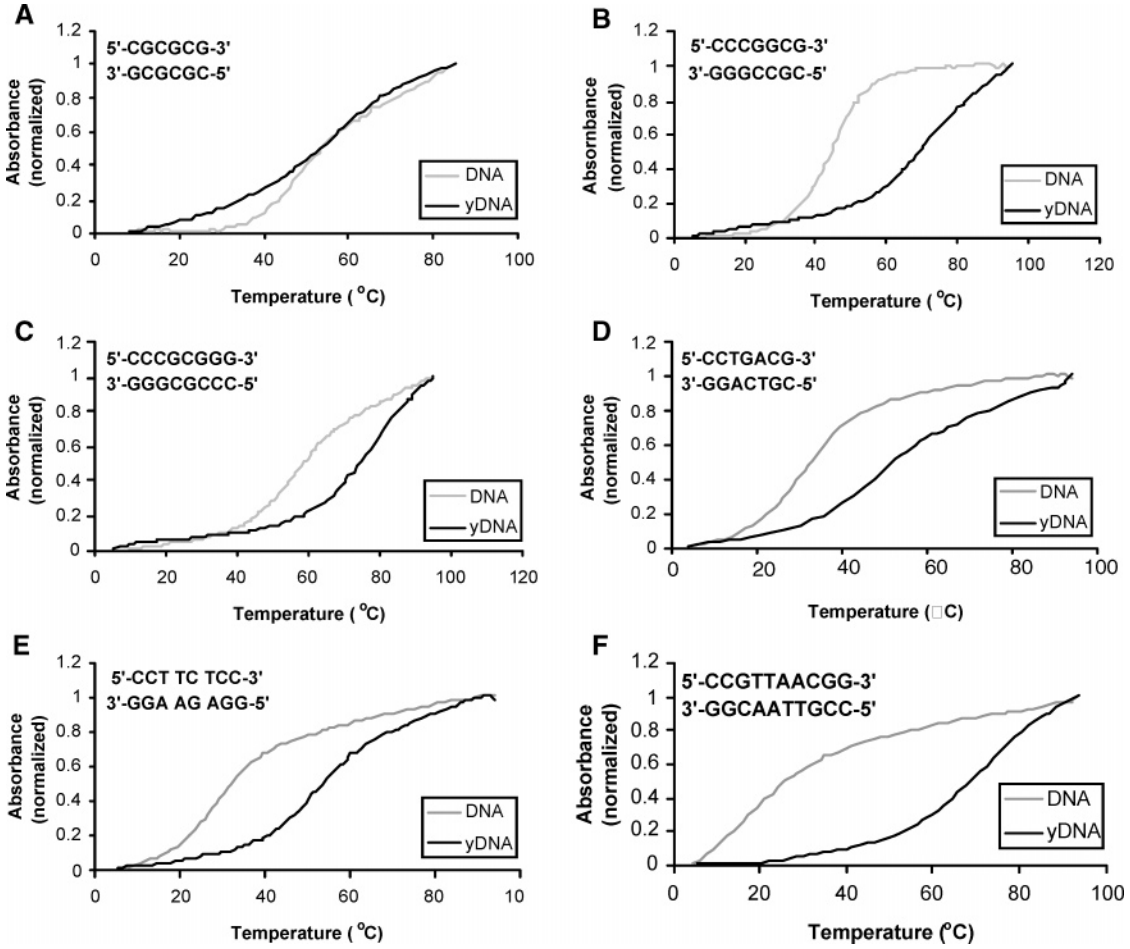


Figure 3. Representative thermal melting profiles (A−F, normalized) for six different sequences of yDNA compared with those of DNAs of the analogous sequence. yDNA curves are in black; DNA controls are in gray. Sequences are as shown. Numerical data are given in Table 1.

yDNA, the same orientation preference that occurs for mixed DNA sequences such as the control sequence (entry 10).

DNA oligonucleotide strands can bind in antiparallel fashion to complementary strands of either DNA or RNA. For mixed sequences there is generally a small preference for binding of RNA, reflecting a greater stability for A-form RNA−DNA helices over B-form DNA duplexes.[13] For yDNA, however, it is not known whether an A-like backbone conformation is possible. Thus, we used this same mixed yT,yC sequence context to evaluate in a preliminary way whether the yDNA oligomer could bind to an RNA complement in antiparallel orientation. The data are given in entries 9 and 18 in Table 2. The results show that the yDNA octamer strand did, in fact, bind the RNA complement, but with considerably lower affinity than the same yDNA strand bound to the DNA complement. The yDNA−RNA complex also was lower in stability than the control DNA bound to the same RNA target. As expected, the natural DNA strand gave a higher-affinity complex with RNA than with DNA, while the results with this sequence showed that the yDNA strand had much higher affinity for a DNA target than for one with an RNA backbone.

Helix Stoichiometry and Ionic Strength Dependence. To evaluate whether yDNA strands have the analogous double-helix-forming behavior that natural DNAs do, it is necessary to confirm that the complexes described above were equimolar, two-stranded helices as is generally observed for DNA. Thus, we carried out strand mixing experiments for three different non-self-complementary sequences taken from Tables 1 and 2. Changes in spectra for the complexes (as compared to each individual strand) were monitored as the mole ratio of the two strands were varied. Plots are given in Supporting Information (Figure S1). Two of the sequences (a heptamer containing yC and a heptamer of mixed sequence) were evaluated by UV absorbance, and showed clear breaks in the plots at mole ratios of 0.42 and 0.54, consistent with 1:1 complexes. The third sequence, a mixed octamer, gave very little change in UV absorption, apparently because of strong single-strand stacking, and therefore we used fluorescence instead to evaluate mixing. Although the change in slope is subtle, the data showed an apparent break in the plot at a mole ratio of 0.53. Taken together, the results give support to the idea that yDNA helices can form 1:1 complexes in a predictable way.

If modified variants of DNA are to be used in practical hybridization experiments, it is important to evaluate how changes in conditions affect the strength of the complexes. Thus, we characterized the ionic strength dependence for a number of yDNA helices, varying $[Na^+]$ over a relatively wide range. Plots of $\log[Na^+]$ versus T_m for three difference sequences are given in Supporting Information (Figure S2), with comparison to natural DNAs of the analogous sequence. The results showed that, in general, the yDNA complexes showed somewhat greater

(13) Chamberlin, M.; Berg, P. *J. Mol. Biol.* **1964**, *8*, 297−313.

Table 2. Thermal Melting Data and Estimated Free Energies for Fully Widened Duplexes Containing yT and/or yC Bases: Effects of Mismatches, Strand Orientation, and Target Backbone[a]

entry	sequence	T_m[b] (°C)	$\Delta G°_{37}$[c] (kcal/mol)
1	5'- ^{y}C ^{y}C ^{y}T ^{y}T ^{y}C ^{y}T ^{y}C ^{y}C$_p$ 3'- $_p$G G A A G A G G	52.3 ± 0.5	-11.1 ± 0.7
2	5'- ^{y}C ^{y}C ^{y}T **T** ^{y}C ^{y}T ^{y}C ^{y}C$_p$ 3'- $_p$G G A **C** G A G G	25.7 ± 0.5	-6.6 ± 0.1
3	5'- ^{y}C ^{y}C ^{y}T **T** ^{y}C ^{y}T ^{y}C ^{y}C$_p$ 3'- $_p$G G A **G** G A G G	38.3 ± 0.5	-8.5 ± 0.1
4	5'- ^{y}C ^{y}C ^{y}T **T** ^{y}C ^{y}T ^{y}C ^{y}C$_p$ 3'- $_p$G G A **T** G A G G	25.6 ± 0.5	-7.0 ± 0.1
5	5'- ^{y}C ^{y}C ^{y}T ^{y}T **^{y}C** ^{y}T ^{y}C ^{y}C$_p$ 3'- $_p$G G A A **A** A G G	nt[d]	nt
6	5'- ^{y}C ^{y}C ^{y}T ^{y}T **^{y}C** ^{y}T ^{y}C ^{y}C$_p$ 3'- $_p$G G A A **C** A G G	nt	nt
7	5'- ^{y}C ^{y}C ^{y}T ^{y}T **^{y}C** ^{y}T ^{y}C ^{y}C$_p$ 3'- $_p$G G A A **T** A G G	nt	nt
8	**5'-** ^{y}C ^{y}C ^{y}T ^{y}T ^{y}C ^{y}T ^{y}C ^{y}C$_p$ **5'-** G G A A G A G G$_p$	nt	nt
9	5'- ^{y}C ^{y}C ^{y}T ^{y}T ^{y}C ^{y}T ^{y}C ^{y}C$_p$[e] 3'-*rG G A A G A G G*	25.0 ± 0.5	-6.6 ± 0.2
10	5'- C C T T C T C C$_p$ 3'- $_p$G G A A G A G G	27.6 ± 0.5	-6.9 ± 0.1
11	5'- C C T **T** C T C C$_p$ 3'- $_p$G G A **C** G A G G	11.0 ± 0.5	-4.1 ± 0.2
12	5'- C C T **T** C T C C$_p$ 3'- $_p$G G A **G** G A G G	18.4 ± 0.5	-5.2 ± 0.4
13	5'- C C T **T** C T C C$_p$ 3'- $_p$G G A **T** G A G G	13.9 ± 0.5	-4.3 ± 0.3
14	5'- C C T T **C** T C C$_p$ 3'- $_p$G G A A **A** A G G	nt	nt
15	5'- C C T T **C** T ^{y}C C$_p$ 3'- $_p$G G A A **C** A G G	nt	nt
16	5'- C C T T **C** T C C$_p$ 3'- $_p$G G A A **T** A G G	nt	nt
17	**5'-** C C T T C T C C$_p$ **5'-$_p$**G G A A G A G G	nt	nt
18	5'- C C T T C T C C$_p$ 3'-*rG G A A G A G G*	41.8 ± 0.5	-9.5 ± 0.1

[a] Conditions: 100 M NaCl, 10 mM $MgCl_2$, and 10 mM PIPES·Na (pH 7.0). [b] T_m values are for 5.0 μM oligonucleotide. [c] Averages of values from van't Hoff and curve fitting methods. [d] "nt" indicates no melting transition > 10 °C. [e] Sequences in italics are RNA.

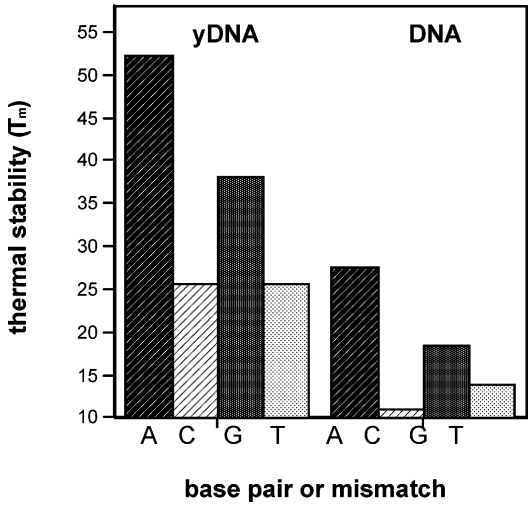


Figure 4. Histogram showing sequence selectivity of yDNA (left) compared with that of DNA (right), as measured by differences in T_m values. The numerical data are given in Table 2. The data are for octamers with a varied central pair involving yT or T opposite each of the four natural bases.

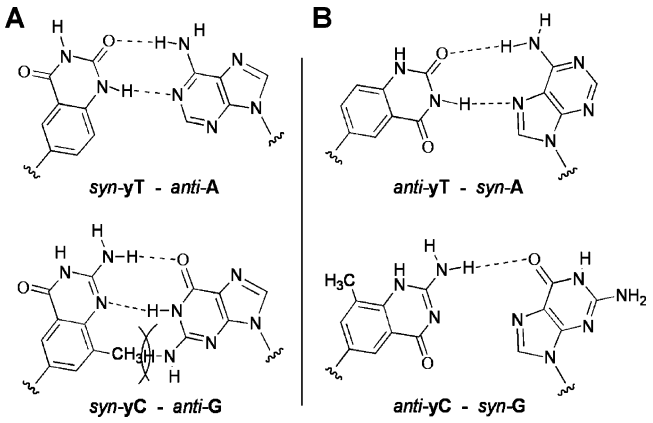


Figure 5. Possible alternative *syn* base pairing structures involving yDNA bases paired opposite their Watson−Crick complements. The favored *anti*,*anti* paired structures are shown in Figure 1. (A) Possible pairs involving *syn* yDNA bases; a steric clash is expected to disfavor this geometry. (B) Possible structures for *syn* DNA bases; this geometry is expected to be disfavored by the energetically unfavorable *syn* purine geometry and by the lack of two H-bonds for the yC−G pair.

ionic strength dependence than the control DNAs did. For example, the heptamer duplex d(CCCGGCG)·d(CGCCGGG) gave a slope of 4.1 °C increase in T_m per log unit increase in sodium concentration, while the yDNA case gave a slope of 5.3 °C. Interestingly, one sequence context, dCCTTCTCC with its complement, showed what appears to be two different dependences on salt, giving a lower slope at high concentrations (>500 mM Na^+) than at normal concentrations. The same dual dependency was seen for the yDNA. This suggests that an alternative conformation (identity currently unknown) for this sequence may be possible at high ionic strength.

Circular Dichroism (CD) Spectra. To begin to collect and evaluate indicators of conformational preferences for yDNAs, we measured CD spectra for six of the sequence contexts in this study. For comparison, we also measured spectra for the natural DNAs of the analogous sequences. The data are shown in Supporting Information (Figure S3). The spectra show variations in shapes of the spectra (both for DNA and yDNA) depending on base composition and sequence. For the most part, the yDNA spectra show a positive peak at 180−210 nm, a negative peak at ∼220−235 nm, and another positive peak at 245−270 nm. This is somewhat analogous to the CD spectra observed previously for yDNA containing the yA base,[6] except that the peaks for yA-containing duplexes are shifted to the red by ca. 20 nm, as the absorption and fluorescence of yA are also red-shifted as compared to that for yT and yC.

Discussion

The current data establish that yDNA strands involving the new benzopyrimidines consistently form highly stable and highly sequence selective helical complexes with their complements. The cases studied by mixing plots show 1:1 binding as would be expected in analogy to natural DNAs, and one experiment demonstrated a clear preference for antiparallel orientation, again analogous to the behavior of DNA. We propose that the local base pair structure of yDNA is likely to

be that shown in Figure 1: namely, hydrogen bonding analogous to Watson−Crick bonds and bases oriented *anti-*,*anti-*. This would yield a helix in which each base pair is widened by 2.4 Å, as designed.

Some caution is warranted in drawing firm conclusions about this base pair structure, however. As is sometimes true for DNA, it is possible that yDNA pairs could adopt other hydrogen bonded geometries, such as *syn*,*anti*. An initial analysis of the most likely alternative base pair geometries (see Figure 5) suggests that they would be less stable than the intended *anti*,*anti* geometry shown in Figure 1. For example, one possibility is that the yDNA bases can flip to a *syn* orientation (Figure 5A), which might still allow for the formation of two hydrogen bonds. However, this seems unlikely since the methyl group of yC would clash severely with the exocyclic amino group of G. Since yC forms highly stable and selective pairs with G, this seems inconsistent with this base pair geometry. A second alternative is that the yDNA bases remain in *anti* orientation, but that the complementary natural bases flip to the *syn* geometry (Figure 5B). This possibility also seems rather unlikely for two reasons: first, it is energetically costly for purines to adopt the *syn* orientation, which is inconsistent with the high stability of yDNA. Second, while a yT−A *anti*,*syn* pair could possibly form two H-bonds, the analogous yC−G pair can only form one. Thus, overall we conclude that the best hypothesis for yDNA structure involves the *anti*,*anti* base pair geometry, but structural studies will be necessary to confirm this.

The expected structures of the yT−A and yC−G pairs involve two and three hydrogen bonds, respectively, which is directly analogous to natural DNA T−A and C−G pairs. Thus, we expect that the stabilization of yDNA from hydrogen bonding is similar in magnitude to the stabilization that such bonds lend to DNA duplexes. For this reason it is likely that the added thermal stability (and presumed thermodynamic stability) of yDNA double helices arises from differences in stacking of the yDNA bases as compared to that of natural DNA bases. An early report of the properties of the yT and yC nucleosides measured their stacking in a dangling-end context and found that yT stacks with a free energy of 0.7 kcal/mol, compared to 0.4 kcal/mol for natural T.[11] Similarly, yC stacks with a free energy of 0.6 kcal/mol, as compared with 0.4 kcal/mol for C. The contribution of stacking to helix stabilization could in principle occur by enhanced stacking in the single-stranded state, which might lower the entropic cost to duplex formation. Alternatively, stacking might be enhanced in the duplex state, which would be reflected in a more favorable enthalpy. The current thermodynamic data are likely too unreliable to draw any conclusions about enthalpy and entropy changes; thus, we do not report $\Delta H°$ and $\Delta S°$ values although they are in principle available from the van't Hoff plots. Experiments are underway to carry out detailed calorimetric measurements to better evaluate the origins of this stability.

The combination of the two new nonnatural widened bases and two natural nucleic acid bases—yT, yC, A, and G—makes a new four-base genetic system encoding the same amount of information in a given sequence length as does natural DNA. Thus, the current yDNA helices make up a second new four-base genetic form, after the previous xDNA system involving xT, xA, T, and A.[1] As with the previous xDNA four-base system, the yDNA system is distinct from the DNA genetic system in its size and stability. The chief difference of the new design is the vector of extension of the putative pairs. The high stability of the examples tested here suggests that the new geometry does not substantially interfere with hydrogen bonding or stacking with adjacent pairs.

The current data show that oligonucleotide strands entirely composed of yDNA nucleosides may be useful as sequence-specific agents for recognition of DNA. For example, a test octameric strand (Table 2) was efficient in binding a natural DNA complement, with a T_m value 24 °C higher and estimated free energy more than 4 kcal/mol more favorable than a natural DNA oligonucleotide. The sequence generality of this binding remains to be seen, however. The current experiments tested this in the context of a yDNA pyrimidine strand recognizing a homopurine sequence. Thus, it is not yet known whether pyrimidine-containing targets could also be complexed in an analogous way by more general yDNA strands. A recent test of a related yA nucleotide showed that it was able to form stable yDNA helices, binding strands containing thymine,[6] but it is not yet known whether yA can be combined with the current yT and yC analogues to form a regular and stable helix. Moreover, for general sequence recognition, an as-yet-unknown yG analogue would also be needed. If these developments can be realized, the high stability and selectivity of yDNA strands, and the inherent fluorescence of yDNA bases,[6,11] might be useful in reporting on sequence recognition.

Experimental Section

Oligonucleotide Synthesis and Characterization. 5′-*O*-Tritylated phosphoramidite derivatives of dyT and dyC were prepared as described in ref 11. Oligonucleotides were synthesized on Applied Biosystems 394 and 392 synthesizers using standard β-cyanoethyl phosphoramidite chemistry. The coupling time for the modified nucleoside phosphoramidites was extended from 100 to 700 s. Stepwise coupling yields for the extended bases were all greater than 98% as determined by trityl cation response. All widened oligonucleotides were synthesized in DMT-off mode on the 5′ end and deprotected from alkyl controlled pore glass (CPG) supports in concentrated ammonia overnight at 55 °C. All oligonucleotides were prepared with 3′ phosphate groups (phosphate-ON reagent, Glen Research) to avoid the need for modified CPG. Oligonucleotides were purified by preparative 20% denaturing polyacrylamide gel electrophoresis (PAGE) and isolated by excision and elution from gel. The recovered material was subsequently quantified by absorption at 260 nm with molar extinction coefficients determined by the nearest neighbor method. All synthesized oligodeoxynucleotides were confirmed by MALDI mass spectroscopy. Data are given in Supporting Information. Oligoribonucleotides were purchased from Dharmacon (Boulder, CO) and were used without further purification.

Thermal Denaturation Studies. Solutions for the thermal denaturation studies contained the indicated concentrations of self-complementary or non-self-complementary strands containing nucleoside analogues. The buffer contained NaCl/$MgCl_2$/Na·PIPES (100 mM:10 mM:10 mM) at pH 7.0. After the solutions were prepared, they were heated to 95 °C and annealed in the UV−vis spectrophotometer Varian Cary 1 equipped with a thermoprogrammer at the rate of 1.0 °C/min. Melting studies were carried out in Teflon-stoppered 1-cm path-length quartz cuvettes under a steady flow of nitrogen current to prevent moisture condensation on quartz wall at low temperature. Absorbance was monitored while the temperature was raised from 5 to 95 °C or lowered from 95 to 5 °C at a rate of 1 °C/min. All sequences were monitored at 260 nm except the marked ones. Computer fitting of the melting data with MeltWin software generatedT_m, free energies (ΔG),

and other thermodynamic data for the duplexes. The curve fit values were compared with those calculated from van't Hoff plots by plotting $1/T_m$ verse $\ln(C_T/4)$ and matched with good agreement.

Circular Dichroism Measurements. CD spectra were recorded on an *AVIV* CD spectrometer (model: 62A DS) from 400 to 180 nm at 15 °C. Data were collected for 1 s in 1-nm intervals in a 1-cm pathlength quartz cell. All samples were in 3 μM total oligonucleotide concentration with 1:1 single strand ratios when paired. The buffer solution contained 100 mM NaCl, 10 mM $MgCl_2$, and 10 mM Na·PIPES, pH = 7.0.

Acknowledgment. This work was supported by the U.S. National Institutes of Health (GM63587). A.H.F.L. was supported by a Croucher Foundation Fellowship.

Supporting Information Available: Data for characterization of oligonucleotides, mixing experiments for measurement of stoichiometry, data for measurement of ionic strength dependence, and CD spectra for six yDNA helices. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0430604