Published in final edited form as:

Chembiochem. 2009 October 12; 10(15): 2530–2538. doi:10.1002/cbic.200900434.

Structure and Replication of yDNA: A Novel Genetic Set Widened by Benzo-homologation

Dr. Haige Lu, Dr. Stephen R. Lynch, Dr. Alex H. F. Lee, and Prof. Dr. Eric T. Kool

Department of Chemistry, Stanford University Stanford, CA 94305-5080 (USA), Fax: (+1)650-725-0259, kool@stanford.edu

Abstract

In a functioning genetic system, the information-encoding molecule must form a regular self-complementary complex (e.g., the base paired double helix of DNA) and it must be able to encode information and pass it on to new generations. Here we study a benzo-widened DNA-like molecule (yDNA) as a candidate for an alternative genetic set, and we explicitly test these two structural and functional requirements. The solution structure of a 10-bp yDNA duplex is measured by 2D-NMR methods for a simple sequence composed of T-yA/yA-T pairs. The data confirm an antiparallel, right-handed, hydrogen-bonded helix resembling B-DNA but with wider diameter and enlarged base pair size. In addition to this, the abilities of two different polymerase enzymes (Klenow fragment of DNA pol I (Kf) and the repair enzyme Dpo4) to synthesize and extend the yDNA pairs T-yA, A-yT, and G-yC are measured by steady-state kinetics studies. Not surprisingly, insertion of complementary bases opposite yDNA bases is inefficient due to the larger base pair size. We find that correct pairing occurs in several cases by both enzymes, but that common and relatively efficient mispairing involving T-yT and T-yC pairs interferes with fully correct formation and extension of pairs by these polymerases. Interestingly, the data show that extension of the large pairs is considerably more efficient with the flexible repair enzyme (Dpo4) than with the more rigid Kf enzyme. The results shed light on the properties of yDNA as a candidate for an alternative genetic information-encoding molecule and as a tool for application in basic science and biomedicine.

Keywords

Wide DNA; genetic set; base pair; polymerase; solution structure; NMR

Introduction

Natural DNA as a genetic molecule serves to store and assist the transfer of biological information, and these functions are enabled by its self-complementary bimolecular structure in a repeating polymer. A basic question that has been of interest to researchers of late is what structural features of DNA are required for this biochemical and biological function, and whether there might exist other structures that can also store and transfer genetic information. Since the DNA bases encode this information, recent studies in multiple laboratories have focused on alternative designs for bases and base pairs that function within the purine-pyrimidine architecture of DNA.[1] Some of these studies have focused on the development of new base pairs to augment the natural set of letters (A,G,C,T) in the genetic alphabet.[2] Such work has led to useful new tools for probing biochemical function[3] and for detecting disease.[4]

Our own laboratory has recently taken a different approach, by asking whether other (non-purine-pyrimidine) architectures for base pairs might also retain such functions. In this regard, we have investigated two designs in which base pairs are enlarged by the addition of a benzene ring (benzo-homologation).[5] One of these has been expanded DNA (xDNA),[6] the design of which was inspired by the lin-benzo-purines first characterized by Leonard.[7] The second is wide DNA (yDNA), the subject of this report, which we developed to have a different vector of benzo-expansion.[8] Three yDNA nucleosides have been prepared and characterized (Fig. 1), potentially yielding a six-letter genetic set. Because of their larger conjugated pi systems, the yDNA bases and base pairs (like xDNA) are predicted to have potentially enhanced properties in charge transfer as compared with natural DNA.[9] Moreover, all of the yDNA and xDNA bases are inherently fluorescent with high quantum yields, suggesting their use in mechanistic and biochemical probes. Finally, both xDNA and yDNA form helical complexes that are more stable than DNA of the analogous sequence. [6a,6e,8a,8c] Although the structure of xDNA has been established by 2D-NMR recently in solution,[6f,6g] the structure of yDNA, with its distinct base pair geometry, has not. Purines with other complementary purines can also lead to bimolecular helices as well.[12] Although none of these larger-than-natural systems has yet been studied structurally by NMR or x-ray crystallography, it is becoming evident that the DNA phosphodiester backbone can readily adapt to base pair architectures that are larger than the natural one. Indeed, evidence has even been presented for base pairing in systems with naphtho-expansion, potentially yielding a 4.8 Å increase over natural base pair size.[13]

In the current study we evaluate two of the important functional properties of yDNA as a potential encoder of genetic information: its structure and its substrate capabilities with DNA polymerase enzymes. In the first NMR structural studies of a yDNA helix, we find that a simple two-letter yDNA sequence can assemble into a base-paired antiparallel double helix that is closely analogous to DNA. Second, we carried out the first detailed studies of the ability of two different DNA polymerases to read sequence encoded by the three yDNA bases. We find that although yDNA pairs can be successfully synthesized and bypassed by such enzymes, there are certain sequence limitations on the success of this replication due to some mismatch geometries that may closely resemble natural base pairs. Overall, the results shed light on structural requirements for designed genetic sets if they are to assemble into helices, and on structural requirements for them to act as substrates for enzymes.

Results

Structure of a yDNA duplex in solution

We undertook studies using 2D-NMR methods to determine the structure of a self-complementary bimolecular complex. Earlier studies had determined that yDNA formed 1:1 complexes that showed cooperative melting behavior, with hypochromicity and CD spectra characteristic of a base-stacked, helical structure.[8a,8c] However, the molecular pictures of base pairing interactions and sugar-phosphate backbone conformation in this complex remained undetermined.

For this first study of yDNA structure we chose a very simple 10-base palindromic, self-complementary sequence composed only of T-yA/yA-T pairs. The full proton assignment of the structure is available in the Supporting Information file. It proved to be relatively well-behaved at 1.3 mM concentration in water and D₂O (20 mM sodium phosphate, pH 7.0, 15 °C (H₂O) or 35 °C (D₂O)), adopting one chief conformation on the NMR timescale. The dispersion of the resonances for the 10 distinct nucleotides enabled the assignment of all protons for the duplex except for the amino protons of yA, and the imino protons of T for the terminal two base-pairs. Thus, the initial data were consistent with antiparallel symmetry

in the bimolecular complex, with the expected hydrogen-bonded pairing between T and yA residues.

The structure of the 10 nucleotide duplex was calculated from 822 observed NOEs, 344 of which were internucleotide, including 36 NOEs from opposing strands. Structural statistics are included in Table S2 for the 21 low energy structures.

The overall structure of this yDNA sequence bears several similarities to B-form DNA (Table 1). The yDNA forms a right-handed, fully base-paired helix, and the backbone conformation is close to that of B-DNA, with a C-2'-*endo* sugar pucker. Base pair orientation and rise are not far from those of natural DNA. The pairing scheme of T with yA involves two hydrogen bonds, closely analogous to those in a natural T-A pair. Figure 2 shows wire-frame structures of the yDNA helix and pairs, and Figure 3 shows space-filling structures of this yDNA duplex compared with B-DNA having the analogous sequence.

The chief differences in the yDNA structure (as compared with natural DNA, Table 1) all stem from the benzo-homologation that widens the base pairs in yDNA. Although the major groove is somewhat narrower, the minor groove of yDNA is considerably wider than that of DNA, and both the overall diameter and the glycosidic bond distance are greater in yDNA than in B-DNA. The depths of the major and minor grooves are greater as well. As a result of the larger helix diameter, it takes more deoxyribosyl phosphodiester units to span a full turn; the consequences of this are a lower helix twist in yDNA (33° per pair) and a greater number of base pairs per turn (11 for yDNA compared with 10 for DNA). Similar effects were seen previously for a different benzo-homolog (xDNA);[6f] a direct comparison of yDNA, xDNA and B-DNA structural parameters are available in the Supporting Information file (Table S1). Finally, the yA base is considerably elongated compared with adenine; as a result, yA bases stack with more surface overlap against other yA residues within the same strand (Fig. 2B); moreover, there is interstrand stacking between yA residues that is not typically present in B-DNA.

Polymerase-mediated synthesis of yDNA pairs

To explore the possibility of polymerase replication of genetic information potentially encoded by yDNA, we carried out single-nucleotide insertion experiments using single yDNA bases in a template strand of DNA. We tested the ability of DNA polymerases to insert natural deoxynucleoside triphosphate derivatives (dNTPs) opposite each of these yDNA bases in separate experiments. We chose two different enzymes for these initial studies: first, the Klenow fragment of *E. coli* DNA polymerase I (exonuclease free (Kf exo-), a well-studied DNA polymerase of the A family, and second, the Dpo4 enzyme from *Sulfolobus solfataricus*, which is a low-fidelity DNA repair enzyme of the Y family.[15] Steric probing experiments have indicated that the former is a relatively rigid enzyme with a strict size preference,[20] while the latter has a more flexible active site in the functional sense.[21] We carried out steady-state kinetics measurements by evaluating degree of radiolabeled primer extension as a function of dNTP concentration; products were separated from the starting primer by gel electrophoresis, and were quantitated by phosphorimaging.

Results of these base pair synthesis studies for the two enzymes are given in Tables 2 and 3 and are plotted in Figure 4. For the Kf exo- enzyme, the kinetics data show that the yA base successfully directs the preferential incorporation of T opposite it, with selectivity ca. 10–100-fold (as judged by $V_{\max}/K_m$ values) for this correct pair over mispairs. However, the enzyme chooses the wrong nucleotide opposite both yT and yC bases, with dTTP being the most efficient substrate opposite these bases. The second best substrate in both cases is the correct one: dATP opposite yT is slightly less efficient than the mispair (by a factor of 3), while dGTP opposite yC is only 2.6-fold less efficient than the T-yC mispairing. This

preferential mispairing of T-yC and T-yT was observed in other cases as well (see below). For the Kf exo- enzyme, the efficiency of synthesis of correct yDNA pairs was low, at 46-fold lower to over 1000-fold lower than synthesis of a normal Watson-Crick base pair.

The results for the Dpo4 enzyme nucleotide insertion experiments are similar (although not identical) to those with the Kf exo-enzyme (compare Tables 2 and 3). Once again, this second enzyme successfully inserted dTTP opposite yA, thus exhibiting “correct” base pair selectivity in that case. However, the Dpo4 enzyme preferred mismatches involving T for the yT and yC bases rather than the correct choices A and G respectively. As with the prior enzyme, the second-best choice for these two cases (lower in efficiency by small factors of 2.2 to 10) was the correct partner (i.e., dATP opposite yT and dGTP opposite yC). The overall efficiency of nucleotide insertion in the correctly matched pairs was lower than that of canonical Watson-Crick pairs by factors of 120–720 depending on the specific pairs. Thus, both of these enzymes can synthesize yDNA pairs at low to moderate efficiency, but both also suffer from interference in the selectivity by relatively efficient T-yT and T-yC mispairing.

Polymerase extension of yDNA pairs

Full synthesis of modified DNA base pairs requires not only that an enzyme assemble a given pair in the first place, but also that this unnatural pair be extended by synthesis of the ensuing pair. In the present template DNA, the next base (beyond the putative yDNA pair) is a G in the template, requiring insertion of dCTP opposite it. We therefore examined the ability of Kf exo- and Dpo4 to extend all the yDNA pairs and mismatches during synthesis of this C-G pair. As above, we carried out steady-state kinetics measurements, the results of which are presented in Tables 4 (Kf exo-) and 5 (Dpo4) (see also Figs. 4C,D).

The data from the extension experiments showed that the Kf enzyme was quite inefficient at extending the correctly matched yDNA pairs, with efficiencies that were lower than natural pairs by large factors of up to 3×10^4 in $V_{\max}/K_m$. The extension selectivity also suffered (as with insertion) from a preference for T mispairing. For example, although the correct yA-T pair was correctly extended with the proper selectivity, when yT or yC were at the termini, the enzyme best extended the mispairs of these bases with T (Table 4).

The Dpo4 enzyme exhibited greater success in the extension of yDNA pairs. With yA in the template, extension proceeded best when it was correctly paired with T (Table 5), and similarly, with yT in the template, extension proceeded most efficiently when it was paired opposite A. Thus two of three pairings functioned correctly. However, the yC-containing terminus yielded the incorrect selectivity in extension, with the yC-T mismatch preferred. Overall, the Dpo4 enzyme was considerably more efficient at extending yDNA pairs than was the Kf exo- enzyme. Extension of the correct yDNA pairs was less efficient than natural pairs by factors of 3-fold (for A-yT) to 810-fold (G-yC) in $V_{\max}/K_m$. This kinetic penalty is considerably less than the above factors of 10^3 – 10^4 that we observed for the Kf enzyme.

Discussion

Taken together, these structural and enzymatic studies shed light on the ability of the yDNA architecture to function correctly in hybridization and in transfer of information via replication. The structural studies show that a simple yDNA sequence is structurally self-consistent and self-complementary, and in that respect is similar both to Watson-Crick B-DNA and to the previously studied xDNA.[6f,6g] The nonstandard base pair geometry in the present case results in a wider diameter, but overall the backbone conformation and base pairing modes are remarkably similar to those of natural DNA. Of course, these conclusions

should be considered not yet general for all yDNA sequences since only one pair has been examined in this initial structure.

Given that caution, the structural data explain the stability of xDNAs containing the yA-T/T-yA pairing. Intra- and inter-strand stacking surface area is enhanced by the larger size of the yA base. Previous studies with modified DNA bases have correlated base size with stacking free energy in DNA,[22] and measurements of the stacking free energy of yA, yC, and yT all indicate that all three of these base stack considerably more favorably than do the analogous natural bases, resulting in substantial double helix stabilization.[8a,8c]

The structural studies also have implications in the general design of alternative base pair architectures, particularly those that are enlarged over purine-pyrimidine systems. Although the ribose-phosphodiester backbone evolved presumably to organize the purine-pyrimidine pair structure, these and other recent studies show that it is flexible enough to adapt readily to base pair structures that are considerably larger. Indeed, the current data show that little change in backbone conformation is needed; previous studies show that this is the case for xDNA as well.[6f] Instead, it appears that an important requirement is that the base pair geometry is *regular* among the different pairings. For example, although yDNA as a whole helix is quite stable, individual yDNA pairs in the context of natural DNA are destabilizing.[8a,8b] Thus it appears that junctions between pairs of different size are destabilizing, while a regular helix of either size is stable.

The finding of a stable, regular paired helix suggests the possible use of yDNAs as probes of DNAs and RNAs, at least in limited applications. Hybridization is robust, and appears to proceed in the predicted way with antiparallel orientation and a conformation close to that of B-DNA. Moreover, the three yDNA bases are inherently fluorescent,[8a,8b] although it is not yet known whether the emission of yDNA bases changes upon hybridization as is the case for xDNA oligomers.[23] However, this application would currently be limited to sequence targets that do not contain C, as there is as yet no yG analog (the analogous yG design is expected to have poor aromaticity[9a]). There are, of course, useful classes of probes where parts of the sequence are general and are not intended to recognize natural DNA, such as in the stems of beacon probes, portions of capture probes, and in other bivalent probes.

The current polymerase studies show that although the Kf *exo*- and Dpo4 enzymes have the capability to make and bypass yDNA pairs when supplied with the correct nucleotides, there are strong limitations on the reading of sequence, due primarily to interference by T-yT pairing. We hypothesize that this T-yT pairing is relatively efficient because it may mimic natural Watson-Crick pairs in geometry, thus fitting the active sites of the polymerases better in the steric and geometric sense than do the intended larger yDNA pairs. Simple examination of structures (Scheme 2) suggests that a T-yT pair has the potential to form two hydrogen bonds with a geometry that places the deoxyribose backbones at a distance and orientation quite close to that of a T-A pair. Structural studies of this mismatch would shed light on this issue. Recent studies have also tested the possibility of cloning yDNA-containing plasmid DNAs in *E. coli*. [24] As in the current polymerase studies, it was found that frequent mispairing of yT with T occurred, thus resulting in plasmids with yT to A mutations after replication had proceeded.[24] Thus the current results help explain those recent findings. Interestingly, the related expanded DNA analog (xDNA) does not have this mispairing issue despite the fact that it is also a benzo-homolog of DNA. Although both yDNA and xDNA are increased in size by addition of benzene rings to DNA bases, they have different orientations for this size expansion.[5] As a result, a T-xT pair is expected to have a geometry similar to that of a “wobble” T-G mismatch (see SI), which makes this mispairing considerably less favorable than it is in yDNA, where it resembles a canonical

base pair. The consequence of this is that base pair synthesis and extension occur with much less mispairing in vitro for the xDNA series.[25]

These results are also consistent with recent studies by Matsuda in the testing of polymerase synthesis of base pairs designed to form four hydrogen bonds. Those studies (which used the Kf exo- enzyme as well) showed a strong polymerase preference for pairs that were the same size as Watson-Crick purine-pyrimidine pairs.[10b] Such a preference is not surprising given that evolution has selected this enzyme to function with the architecture of natural DNA.

The current yDNA design as a self-contained genetic set contains six letters (two more than the natural set). The present results show that attempts to replicate this full set using natural polymerase enzymes, with their bias toward purine-pyrimidine base pair size, may result in frequent mutation by mispairing if all six letters are included. However, the data do suggest the interesting possibility that a subset of these six might be more successful. For example, the omission of yA and T from the set (leaving yC, yT, A and G) would avoid the issue of T mispairing, thus possibly allowing the storage and transmission of information with higher fidelity. Further work will be needed to investigate this possibility.

Experimental Section

Nucleoside Phosphoramidite Derivatives of dyA, dyC, dyT

The syntheses of these three compounds were carried out as previously reported.[8a,8b]

Oligonucleotide Synthesis

Oligodeoxynucleotides (28mer template DNAs and the 10mer for NMR studies) were synthesized on an Applied Biosystems 394 DNA/RNA synthesizer on a 1 mmole scale. Coupling employed standard β -cyanoethyl phosphoramidite chemistry, but with extended coupling time (600 s) for nonnatural nucleotides. The oligomers were deprotected in concentrated ammonium hydroxide (55 °C, 16 h), purified by preparative 20% denaturing polyacrylamide gel electrophoresis, and isolated by excision and extraction from the gel, followed by dialysis against water. The recovered material was quantified by absorbance at 260 nm with molar extinction coefficients determined by the nearest neighbor method. Molar extinction coefficients for unnatural oligomers were estimated by adding the measured value of the molar extinction coefficient of the unnatural nucleoside (at 260nm) to the calculated value for the natural DNA fragments. Previous studies have shown that yDNA bases have very low hypochromicity in yDNA oligomers.[8a,8c] Molar extinction coefficients used for yDNA nucleosides were as follows: dyA, $\epsilon_{260}=57,300 \text{ M}^{-1}\cdot\text{cm}^{-1}$; dyT, $\epsilon_{260}=1,200 \text{ M}^{-1}\cdot\text{cm}^{-1}$; dyC, $\epsilon_{260}=5,800 \text{ M}^{-1}\cdot\text{cm}^{-1}$. The oligomers containing these three bases were characterized by MALDI-TOF mass spectrometry (data are given in the Supporting Information file).

Polymerase kinetics methods

Kf polymerase—The 5'-terminus of the primer was labeled using [γ - ^{32}P]ATP and T4 polynucleotide kinase. The labeled primer was annealed to the template in a buffer of 100 mM Tris•HCl (pH 7.5), 20 mM MgCl_2 , 2 mM DTT, and 0.1 mg/mL acetylated BSA. Polymerase reactions were started by mixing equal volumes of solution A containing the DNA-enzyme complex and solution B containing dNTP substrates and terminated by adding 1.5 volumes of stop buffer (95% formamide, 20 mM EDTA, 0.05% xylene cyanol and bromophenol blue). Steady state kinetics measurements for standing start single nucleotide insertions were carried out as described.[14] The final duplex DNA concentration was 5 or 0.5 μM . Final concentrations of triphosphates (0.5–500 μM), amount of polymerase used

(0.005–0.1 u/μL) and reaction time (1–60 min) were adjusted to give <20% conversion. Extents of reaction were determined by running quenched reaction samples on 20% denaturing polyacrylamide gel. Relative velocities were calculated as product concentration divided by reaction time and normalized to 0.005u/μL enzyme concentration.

Dpo4 polymerase—*E. coli* expressing Dpo4 was a gift from Dr. R. Woodgate (National Institutes of Health). The protein was purified according to the published method.[15] The concentration was quantitated by the Bradford method. Primer 5' termini were labeled using [γ - 32 P]ATP (Amersham Bioscience) and T4 polynucleotide kinase (Invitrogen). The labeled primer (~250 nM), template (100 μM) and unlabeled primer (100 μM) were mixed in 2x reaction buffer and gave a total concentration of primer–template of 20 μM. The reaction buffer (1x) contained 40 mM Tris–HCl (pH 8.0), 5 mM MgCl₂, 10 mM dithiothreitol, 250 μg/mL bovine serum albumin (BSA), 2.5% glycerol. The primer–template duplexes were annealed by heating to 90°C, and cooling slowly to 4 °C over 1 hour. The annealed duplex DNA solution (2.5 μL) was mixed with Dpo4 (2.5 μL), and the reaction was initiated by adding a solution of the appropriate dNTP (5 μL). Dpo4 concentration (0.01–1.0 μM) and reaction time (2–60 min) were adjusted with different dNTP concentration (1–500 μM) to give less than 20% incorporation. Reactions were quenched with 15 μL of 95% formamide/10 mM EDTA containing 0.05% xylene cyanol and 0.05% bromophenol blue. Extents of reaction were determined by running quenched reaction samples on a 20% polyacrylamide/7 M urea gel. Radioactivity was quantified using a Phosphorimager (Molecular Dynamics) and the ImageQuant Program. Reaction velocity v (M·min^{−1}) was defined as $v = [S] \cdot I_{\text{ext}} / [(I_{\text{prim}} + I_{\text{ext}}) \cdot t]$, where $[S]$ is the concentration of duplex, I_{ext} and I_{prim} are the intensities of the extended product and the remaining primer, respectively. The V_{max} and K_{m} values were obtained from Hanes-Woolf plots.

NMR methods

Sample Preparation—The yA phosphoramidite was synthesized and incorporated into a self-complementary DNA-like 10mer, 5'-d(yATyAyATyATTyAT), using an Applied Biosystems 394 DNA/RNA Synthesizer. After the oligonucleotide was synthesized without trityl protection on 5'-OH, it was purified by preparative denaturing polyacrylamide gel electrophoresis (PAGE). The purity was found to be greater than 95% as checked by analytical PAGE. The oligonucleotide was desalted by extensive dialysis washing, followed by further centrifugal filtration with Microcon (Millipore Co., Bedford, MA 01730, USA), YM3000 (NMWC=3000). The sample was dissolved in buffer (10 mM Sodium Phosphate, pH 7.0) at a final concentration of 1.3 mM.

NMR spectroscopy—NMR experiments were acquired on either a Varian Inova 600 MHz NMR instrument or Varian NMR System 400 MHz NMR instrument equipped triple resonance and z-gradient capabilities. NMR spectra were acquired at 35°C in 20 mM Sodium Phosphate, pH 7.0, exchanged into 99.996% D₂O, or at 25°C in 90% H₂O/10% D₂O in 20 mM Sodium Phosphate buffer, pH 7.0. ¹H resonances were assigned with standard methods using a combination of DQF-COSY, TOCSY with mixing times of 20 ms and 80 ms, ¹H/¹³C HSQC, ¹H/³¹P Heteronuclear COSY, NOESY, and SS-NOESY experiments. The spectral width in the ¹H dimension for all experiments in D₂O was 10 ppm. ¹H/¹³C HSQC was acquired with 1024 points in t₂ dimension by 128 real points in t₁ dimension with a recycle delay of 2.5 seconds and a carbon spectral width of 45 ppm for the aromatic resonances, and 1024 by 128 with 1.0 seconds recycle delay and a spectral width of 45 ppm. DQF-COSY was acquired with 16 scans of 8192 points in t₂ dimension by 512 points in t₁ dimension, and a recycle delay of 0.75 s. TOCSY experiments were acquired with 16 scans of 4096 points in t₂ by 256 points in t₁, a recycle delay of 0.75 seconds, and either 20 ms or 80 ms mixing times. NOESY experiments in D₂O were acquired with

mixing times of 50 ms, 100 ms, 150 ms, and 250 ms with 48 scans of 4096 points in t_2 and 350 in t_1 , and a recycle delay of 1.95 s. SS-NOESY experiments were acquired in 90% H_2O with mixing times of 75 ms and 250 ms, with 64 scans of 4096 by 256 points, and a recycle delay of 1.5 seconds 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, Sparky 3.1.[16]

The protons on the deoxyribose residues were distinguished by type with $^1H/^{13}C$ 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 with strong intranucleotide NOEs. The protons on the aromatic ring could be distinguished from each other by ^{13}C chemical shift as observed in $^1H/^{13}C$ HSQC. These proton resonances could be assigned by a combination of TOCSY correlation to each other through-bond or NOESY correlation through-space. Sequential assignments could be made through many NOEs observed between sequential nucleotides including protons on the deoxyribose to the protons on the sequential base and protons on one base to protons on the sequential base. The imino protons of T were assigned based upon imino-imino NOEs between sequential base-pairs, and intense NOEs between the T imino to the H2 proton of yA. Additional NOEs were observed between the H2 of yA imino protons of sequential base-pairs to confirm assignments.

Structure Calculation—Structures of the 10 nucleotide DNA duplex were calculated on an Apple MacBook running OS 10.5 with restrained molecular dynamics followed by energy minimization with the program XPLOR-NIH.[17] 100 random starting structures were subjected to a simulated annealing protocol with restrained molecular dynamics utilizing 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 $\text{\AA}^2$; torsion angle force constants were varied from 10 to 50 kcal/mol rad^2 . The high temperature annealing was done with 15000 cycles molecular dynamics at 1000K with low values of interatomic repulsion, followed by 1000 steps of dynamics while increasing the van der Waals force constant, and slow cooling with 500 steps of dynamics while cooling to 300K, increasing the dihedral force constant from 10 to 50 kcal/mol rad^2 . An initial energy minimization with 1000 steps with both the NOE and dihedral force constants at 50 kcal/mol rad^2 was accomplished. Of 100 initial structures, the 21 lowest energy structures were subjected to a second stage of 1000 cycles of molecular dynamics at 1000K and dynamics while cooling to 300K. This step was followed by a final molecular dynamics at 300K while increasing the force constant on the van der Waals potentials and a final energy minimization that included all of the above constraints in addition to attractive Lennard-Jones potentials without electrostatic potentials for 50000 steps. The final structures were superimposed and displayed with the program MacPyMOL.[18]

Distance restraints for non-exchangeable protons were assigned based upon analysis of crosspeak intensity in D_2O NOESY experiments with mixing times of 50 ms, 100 ms, 150 ms and 300 ms. Distance restraints were given ranges of 1.7–2.5 $\text{\AA}$ or 1.7–3.0 $\text{\AA}$ for NOEs observed in 50 ms by comparison to NOEs observed for the known distances for the 2 DNA bases. NOEs observed at 100 ms were assigned a range of 2.5–4.5 $\text{\AA}$. NOEs observed at 150 ms were assigned a range of 3.0–6.0 $\text{\AA}$, and those that were only observed at 300 ms were assigned a range of 3.5–7.5 $\text{\AA}$. NOESY experiment by comparison to known distances from the 2 bases. Hydrogen bonds for the Watson-Crick like base pairs were assigned based upon

intense NOEs observed between the thymine imino proton and y-adenosine H2 as well as the downfield chemical shift of the imino proton resonance.

Dihedral constraints were assigned based on analysis of DQF-COSY and $^1\text{H}/^{31}\text{P}$ heteronuclear COSY. The sugar pucker was constrained based upon comparison of $^3\text{J}_{\text{H1}'\text{H2}'}$, $^3\text{J}_{\text{H1}'\text{H2}''}$, and $^3\text{J}_{\text{H3}'\text{H4}'}$ values. The backbone torsion angle b was estimated based upon measurement of $^3\text{J}_{\text{H5}'\text{P}}$, $^3\text{J}_{\text{H5}''\text{P}}$, and $^4\text{J}_{\text{H4}'\text{P}}$, and assigned a range of $\pm 40^\circ$. The backbone torsion angle g was estimated based upon measurement of $^3\text{J}_{\text{H4}'\text{H5}'}$, $^3\text{J}_{\text{H4}'\text{H5}''}$, and $^4\text{J}_{\text{H4}'\text{P}}$, and assigned a range of $\pm 40^\circ$. The backbone torsion angle e was estimated based upon measurement of $^3\text{J}_{\text{H3}'\text{P}}$, and assigned a range of $\pm 60^\circ$.

Helical Parameters—The helical parameters and torsion angles for the 10 nucleotide duplex were determined by inputting the structures after final energy minimization in pdb format into the program Curves 5.3.[19]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the U.S. National Institutes of Health (GM63587). H.L. acknowledges an Abbott Laboratories Graduate Fellowship, and A.H.-F.L. acknowledges a postdoctoral fellowship from the Croucher Foundation. We thank Dr. Andrew Krueger for obtaining mass spectrometry data.

References

1. a) Switzer C, Moroney SE, Benner SA. J Am Chem Soc. 1989; 111:8322–8323. b) Tor Y, Dervan PB. J Am Chem Soc. 1993; 115:4461–4467. c) Schweitzer BA, Kool ET. J Am Chem Soc. 1995; 117:1863–1872. [PubMed: 20882111] d) Moran S, Ren RXF, Rumney S, Kool ET. J Am Chem Soc. 1997; 119:2056–2057. [PubMed: 20737028] e) Rappaport HP. Nucleic Acids Res. 1988; 16:7253–7267. [PubMed: 3412886] f) Matray TJ, Kool ET. J Am Chem Soc. 1998; 120:6191–6192. [PubMed: 20852721] g) McMinn DL, Ogawa AK, Wu YQ, Liu JQ, Schultz PG, Romesberg FE. J Am Chem Soc. 1999; 121:11585–11586. h) Mitsui T, Kitamura A, Kimoto M, To T, Sato A, Hirao I, Yokoyama S. J Am Chem Soc. 2003; 125:5298–5307. [PubMed: 12720441] i) Paul N, Nashine VC, Hoops G, Zhang P, Zhou J, Bergstrom DE, Davisson VJ. Chem Biol. 2003; 10:815–825. [PubMed: 14522052] j) Kool ET. Accounts Chem Res. 2002; 35:936–943. k) Krueger AT, Kool ET. Chem Biol. 2009; 16:242–248. [PubMed: 19318205]
2. a) Benner SA, Battersby TR, Eschgfäller B, Hutter D, Kodra JT, Lutz S, Arslan T, Baschlin DK, Blattler M, Egli M, Hammer C, Held HA, Horlacher J, Huang Z, Hyrup B, Jenny TF, Jurczyk SC, König M, von Krosigk U, Lutz MJ, MacPherson LJ, Moroney SE, Muller E, Nambiar KP, Piccirilli JA, Switzer CY, Vogel JJ, Richert C, Roughton AL, Schmidt J, Schneider KC, Stackhouse J. Pure Appl Chem. 1998; 70:263–266. [PubMed: 11542721] b) Henry AA, Romesberg FE. Curr Opin Chem Biol. 2003; 7:727–733. [PubMed: 14644182] c) Hirao I. Curr Opin Chem Biol. 2006; 10:622–627. [PubMed: 17035074]
3. a) Schweitzer BA, Kool ET. J Org Chem. 1994; 59:7238–7242. [PubMed: 20882116] b) Moran S, Ren RXF, Kool ET. Proc Natl Acad Sci USA. 1997; 94:10506–10511. [PubMed: 9380669] c) Lan T, McLaughlin LW. Biochemistry. 2001; 40:968–976. [PubMed: 11170418] d) Parsch J, Engels JW. Nucleosides Nucleotides Nucleic Acids. 2001; 20:815–818. [PubMed: 11563122] e) Kincaid K, Beckman J, Zivkovic A, Halcomb RL, Engels JW, Kuchta RD. Nucleic Acids Res. 2005; 33:2620–2628. [PubMed: 15879351] f) Zhang X, Lee I, Berdis AJ. Biochemistry. 2005; 44:13101–13110. [PubMed: 16185078]
4. a) Collins ML, Irvine B, Tyner D, Fine E, Zayati C, Chang C, Horn T, Ahle D, Detmer J, Shen LP, Kolberg J, Bushnell S, Urdea MS, Ho DD. Nucleic Acids Res. 1997; 25:2979–2984. [PubMed: 9224596] b) Paris PL, Langenhan J, Kool ET. Nucleic Acids Res. 1998; 26:3789–3793. [PubMed:

- 9685497] c) Matray TJ, Kool ET. *Nature*. 1999; 399:704–708. [PubMed: 10385125] d) Prudent JR. *Expert Rev Mol Diagn*. 2006; 6:245–252. [PubMed: 16512783]
5. Krueger AT, Lu H, Lee AHF, Kool ET. *Accts Chem Rev*. 2007; 40:141–150.
6. a) Liu H, Gao J, Lynch S, Maynard L, Saito D, Kool ET. *Science*. 2003; 302:868–871. [PubMed: 14593180] b) Liu H, Gao J, Saito D, Maynard L, Kool ET. *J Am Chem Soc*. 2004; 126:1102–1109. [PubMed: 14746479] c) Liu H, Gao J, Kool ET. *J Org Chem*. 2004; 70:639–647. [PubMed: 15651812] d) Liu H, Gao J, Kool ET. *J Am Chem Soc*. 2005; 127:1396–1402. [PubMed: 15686371] e) Gao J, Liu H, Kool ET. *Angew Chem Int Ed*. 2005; 44:3118–3122. f) Liu H, Lynch SR, Kool ET. *J Am Chem Soc*. 2004; 126:6900–6905. [PubMed: 15174859] g) Lynch SR, Liu H, Gao J, Kool ET. *J Am Chem Soc*. 2006; 128:14704–14711. [PubMed: 17090058]
7. Leonard NJ. *Acc Chem Res*. 1982; 15:128–135.
8. a) Lu H, He K, Kool ET. *Angew Chem Int Ed*. 2004; 43:5834–5836. b) Lee AHF, Kool ET. *J Org Chem*. 2004; 70:132–140. [PubMed: 15624915] c) Lee AHF, Kool ET. *J Am Chem Soc*. 2005; 127:3332–3338. [PubMed: 15755149]
9. a) Fuentes-Cabrera M, Sumpter BG, Lipkowski P, Wells JC. *J Phys Chem B*. 2006; 110:6379–6384. [PubMed: 16553457] b) Lait LA, Rutledge LR, Millen AL, Wetmore SD. *J Phys Chem B*. 2008; 112:12526–12536. [PubMed: 18771305] c) Zhang L, Li H, Chen X, Cukier RI, Bu Y. *J Phys Chem B*. 2009; 113:1173–1181. [PubMed: 19159339]
10. a) Hikishima S, Minakawa N, Kuramoto K, Fujisawa Y, Ogawa M, Matsuda A. *Angew Chem Int Ed*. 2005; 44:596–598. b) Minakawa N, Ogata S, Takahashi M, Matsuda A. *J Am Chem Soc*. 2009; 131:1644–1645. [PubMed: 19146369]
11. Doi Y, Chiba J, Morikawa T, Inouye MJ. *J Am Chem Soc*. 2008; 130:8762–8768. [PubMed: 18543918]
12. a) Evertsz EM, Rippe K, Jovin TM. *Nucleic Acids Res*. 1994; 22:3293–3303. [PubMed: 8078763] b) Battersby TR, Albalos M, Friesenhahn MJ. *Chem Biol*. 2007; 14:525–531. [PubMed: 17524983] c) Heuberger BD, Switzer C. *ChemBiochem*. 2008; 9:2779–2783. [PubMed: 18985646]
13. Lee AHF, Kool ET. *J Am Chem Soc*. 2006; 128:9219–9230. [PubMed: 16834396]
14. Goodman MF, Creighton S, Bloom LB, Petruska J. *Crit Rev Biochem Mol Biol*. 1993; 28:83–126. [PubMed: 8485987]
15. Boudsocq F, Iwai S, Hanaoka F, Woodgate R. *Nucleic Acids Res*. 2001; 29:4607–4616. [PubMed: 11713310]
16. Goddard TD, Huang CC, Ferrin TE. *Structure*. 2005; 13:473–482. [PubMed: 15766548]
17. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. *J Magn Res*. 2003; 160:66–74.
18. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; Palo Alto, CA USA: 2008. <http://www.pymol.org>
19. Lavery R, Sklenar H. *J Biomol Struct Dyn*. 1988; 6:63–91. [PubMed: 2482765]
20. Kim TW, Delaney JC, Essigmann JM, Kool ET. *Proc Natl Acad Sci USA*. 2005; 102:15803–15808. [PubMed: 16249340]
21. Mizukami S, Kim TW, Helquist SA, Kool ET. *Biochemistry*. 2006; 45:2772–2778. [PubMed: 16503632]
22. Guckian KM, Schweitzer BA, Ren RXF, Sheils CJ, Tahmassebi DC, Kool ET. *J Am Chem Soc*. 2000; 122:2213–2222. [PubMed: 20865137]
23. Krueger AT, Kool ET. *J Am Chem Soc*. 2008; 130:3989–3999. [PubMed: 18311973]
24. Chelliserrykattil C, Lu H, Lee AHF, Kool ET. *ChemBioChem*. 2008; 9:2976–2980. [PubMed: 19053129]
25. Lu H, Krueger AT, Gao J, Liu H, Kool ET. *J Am Chem Soc*. 2009 submitted.

Figure 2.

Structure of yDNA helix and base pairs. A. Views of structure from major groove (left) and minor groove (right). B. Structures of three consecutive base pairs in the yDNA structure, showing pair geometry and base-base overlap. Shown are T2-yA19, yA3-T18, and yA4-T17. The middle pair is highlighted for clarity. C. Numbering of residues in the helix.

Figure 3. Space-filling models of yDNA helix (blue) compared with B-DNA of the same sequence (red). A. Major groove view; B. Minor groove view.

Figure 4.

Plots comparing kinetic efficiencies and selectivities for yDNA pair synthesis and extension for Kf exo- and Dpo4 enzymes. A) Nucleotide insertion by Kf exo-; B) nucleotide insertion by Dpo4; C) primer/template base pair extension by Kf exo- (extending various pairs/ mismatches shown); D) base pair extension by Dpo4 (* = estimated maximum value; see Tables 3–5).

Scheme 1.

Nucleosides and base pairs used in this study. (A) The three yDNA monomer nucleosides. (B) Putative hydrogen-bonded structures of two yDNA pairs. The current study confirms the yA-T structure.

Scheme 2.

Structures illustrating hypothesized geometric similarity of the common T-yT mispair with that of a normal Watson-Crick T-A pair. Note similar locations of deoxyribose backbone (R) in both cases. See SI for additional structures.

Table 1

Average Structural Parameters for yDNA and B-DNA

parameter	yDNA	B-DNA
Helix Handedness	Right	Right
bp/repeating unit	1	1
bp/turn	11	10
Helix twist (°)	33	36
Rise/bp (Å)	3.2	3.4
Helix pitch (°)	34	34
Base pair inclination (°)	5.9	2.4
Diameter (Å)	20.3	18.4
C1' – C1' distance	12.8	10.7
X displacement from bp to helix axis (Å)	−1.1	−0.2
Glycosidic bond orientation	anti	anti
Sugar conformation	2'-endo	2'-endo
Major groove depth (Å)	7.0	4.1
Major groove width (Å)	8.6	11.4
Minor groove depth (Å)	5.7	5.5
Minor groove width (Å)	7.1	5.9

Table 2Steady-state kinetics data for Kf (exo-)-catalyzed insertion of natural dNTPs opposite yDNA template bases.^a

Template Base	dNTP	$V_{\max}$ ($\mu\text{M}\cdot\text{min}^{-1}$)	K_m (μM)	$10^6 \times V_{\max}/K_m$ (min^{-1})
YA	A	0.0026 $\pm$ 0.0002	118 $\pm$ 14	2.2 $\times$ 10 ¹
YA	C	0.0020 $\pm$ 0.0004	117 $\pm$ 29	1.8 $\times$ 10 ¹
YA	G	0.00016 $\pm$ 0.00001	110 $\pm$ 10	1.5 $\times$ 10 ⁰
YA	T	0.055 $\pm$ 0.005	290 $\pm$ 37	2.0 $\times$ 10 ²
YC	A	0.007 $\pm$ 0.001	106 $\pm$ 13	6.5 $\times$ 10 ¹
YC	C	0.047 $\pm$ 0.012	213 $\pm$ 16	2.3 $\times$ 10 ²
YC	G	0.125 $\pm$ 0.045	72 $\pm$ 9	1.7 $\times$ 10 ³
YC	T	0.425 $\pm$ 0.115	98 $\pm$ 5	4.3 $\times$ 10 ³
YT	A	0.44 $\pm$ 0.09	42 $\pm$ 10	1.2 $\times$ 10 ⁴
YT	C	0.155 $\pm$ 0.035	53 $\pm$ 3	2.9 $\times$ 10 ³
YT	G	0.029 $\pm$ 0.003	85 $\pm$ 5	3.4 $\times$ 10 ²
YT	T	0.31 $\pm$ 0.08	8.3 $\pm$ 0.4	3.6 $\times$ 10 ⁴
T	A	1.8 $\pm$ 0.2	3.4 $\pm$ 1.4	5.5 $\times$ 10 ⁵
C	G	0.8 $\pm$ 0.3	0.6 $\pm$ 0.3	1.4 $\times$ 10 ⁶

^aConditions: 5 μM template-primer duplex, 100 mM Tris $\cdot$ HCl (pH=7.5), 10 mM MgCl₂, 1 mM dithiothreitol and 100 $\mu\text{g}/\text{mL}$ bovine serum albumin, various concentrations of dNTPs, incubated at 37 °C in a reaction volume of 10 μL . Data are averaged from triplicates; standard deviations are shown. $V_{\max}$ is normalized to 0.005 unit/ μL enzyme concentration.

Table 3Steady-state kinetics for insertion of single nucleotides opposite yDNA bases by Dpo4 polymerase. ^a

Template Base	dNTP	V _{max} (μM·min ⁻¹)	K _m (μM)	10 ⁶ ×V _{max} /K _m (min ⁻¹)
YA	A	0.030±0.003	488±26	6.1×10 ¹
YA	C	0.018±0.004	270±130	9.3×10 ¹
YA	G	ND ^b	ND ^b	<10 ¹
YA	T	0.24±0.07	650±140	3.6×10 ²
YC	A	0.026±0.005	237±52	1.1×10 ²
YC	C	0.035±0.010	213±55	1.7×10 ²
YC	G	0.42±0.04	500±190	1.1×10 ³
YC	T	2.6±0.7	234±39	1.1×10 ⁴
YT	A	1.1±0.1	256±31	4.5×10 ³
yT	C	0.44±0.11	230±25	1.9×10 ³
yT	G	0.10±0.01	542±59	1.9×10 ²
yT	T	2.8±0.8	278±52	9.9×10 ³
A	T	4.4±0.3	18±3	2.6×10 ⁵
C	G	5.5±0.5	21±1	2.6×10 ⁵
G	C	11±3	8±1	1.3×10 ⁶
T	A	7.9±0.7	15±2	5.4×10 ⁵
A	A	0.15±0.02	348±81	4.5×10 ²
G	T	0.28±0.12	330±160	9.8×10 ²
T	G	0.77±0.03	385±30	2.0×10 ³

^a Conditions: 5 μM template-primer duplex, 40 mM Tris-HCl (pH=8.0), 5 mM MgCl₂, 10 mM dithiothreitol, 250 μg/mL bovine serum albumin, and various concentrations of dNTPs, incubated at 37 °C in a reaction volume of 10 μL. Data are averaged from triplicates; standard deviations are shown. V_{max} is normalized to 1 μM enzyme concentration.

^b Not determined due to very low insertion efficiency.

Table 4

Steady-state kinetics data for Kf (exo-)-catalyzed extension of yDNA pairs and mismatches during synthesis of a C-G pair immediately beyond them.^a

Templat E Base	Primer Base	$V_{\max}$ ($\mu\text{M}\cdot\text{min}^{-1}$)	K_m (μM)	$10^6 \times V_{\max}/K_m$ (min^{-1})
yA	A	0.0012 $\pm$ 0.0001	144 $\pm$ 9	8.2 $\times 10^0$
yA	C	0.00032 $\pm$ 0.00003	137 $\pm$ 20	2.4 $\times 10^0$
yA	G	ND ^b	ND ^b	<10 ⁰
yA	T	0.053 $\pm$ 0.0010	133 $\pm$ 3	4.0 $\times 10^1$
yC	A	ND ^b	ND ^b	<10 ⁰
yC	C	0.020 $\pm$ 0.003	133 $\pm$ 14	1.5 $\times 10^2$
yC	G	ND ^b	ND ^b	<10 ⁰
yC	T	0.021 $\pm$ 0.05	36 $\pm$ 3	5.8 $\times 10^3$
yT	A	0.11 $\pm$ 0.04	199 $\pm$ 39	5.4 $\times 10^2$
yT	C	0.066 $\pm$ 0.010	99 $\pm$ 20	6.7 $\times 10^2$
yT	G	0.00021 $\pm$ 0.00008	135 $\pm$ 9	1.6 $\times 10^0$
yT	T	0.28 $\pm$ 0.07	7.1 $\pm$ 0.4	3.9 $\times 10^4$
A	T	0.44 $\pm$ 0.05	0.42 $\pm$ 0.09	1.2 $\times 10^6$
C	G	0.30 $\pm$ 0.06	0.31 $\pm$ 0.05	9.6 $\times 10^5$
G	C	0.21 $\pm$ 0.03	0.45 $\pm$ 0.19	6.4 $\times 10^5$
T	A	0.24 $\pm$ 0.06	0.20 $\pm$ 0.11	2.4 $\times 10^6$
A	A	0.0025 $\pm$ 0.0003	122 $\pm$ 13	2.1 $\times 10^1$
G	T	0.084 $\pm$ 0.030	300 $\pm$ 110	2.8 $\times 10^2$

^a Conditions: 0.5 μM template-primer duplex, 100 mM Tris $\cdot$ HCl (pH=7.5), 10 mM MgCl₂, 1 mM dithiothreitol and 100 $\mu\text{g}/\text{mL}$ bovine serum albumin, various concentrations of dCTP, incubated at 37 °C in a reaction volume of 10 μL . Data are averaged from triplicates; standard deviations are shown. $V_{\max}$ is normalized to 0.005 unit/ μL enzyme concentration.

^b Not determined due to low extension efficiency.

Table 5

Steady-state kinetics data for Dpo4-catalyzed extension of yDNA pairs and mismatches during synthesis of a C-G pair immediately beyond them.^a

Template Base	dNTP	$V_{\max}$ ($\mu\text{M}\cdot\text{min}^{-1}$)	K_m (μM)	$10^6 \times V_{\max}/K_m$ (min^{-1})
yA	A	ND ^b	ND ^b	$<10^3$
yA	C	ND ^b	ND ^b	$<10^3$
yA	G	ND ^b	ND ^b	$<10^3$
yA	T	1.2 ± 0.3	73 ± 18	1.5×10^4
yC	A	0.25 ± 0.02	46 ± 9	5.7×10^3
yC	C	0.68 ± 0.04	62 ± 3	1.1×10^4
yC	G	0.71 ± 0.20	146 ± 43	5.2×10^3
yC	T	4.3 ± 0.4	14 ± 2	3.2×10^5
yT	A	16.5 ± 1.8	11 ± 2	1.6×10^6
yT	C	ND ^b	ND ^b	$<10^5$
yT	G	ND ^b	ND ^b	$<10^5$
yT	T	18.1 ± 2.4	36 ± 13	5.9×10^5
A	T	18 ± 4	4.5 ± 0.7	4.0×10^6
C	G	31 ± 4	7.2 ± 0.1	4.2×10^6
G	C	23 ± 4	14 ± 4	1.6×10^6
T	A	34 ± 3	6.2 ± 0.6	5.5×10^6
G	T	2.9 ± 0.2	140 ± 23	2.2×10^4
T	G	6.0 ± 0.5	156 ± 14	3.9×10^4

^a Conditions: 5 μM template-primer duplex, 40 mM Tris-HCl (pH=8.0), 5 mM MgCl_2 , 10 mM dithiothreitol, 250 $\mu\text{g/mL}$ bovine serum albumin, and various concentrations of dCTP, incubated at 37 °C in a reaction volume of 10 μL . Data are averaged from triplicates; standard deviations are shown. $V_{\max}$ is normalized to 1 μM enzyme concentration.

^b Not determined due to very low insertion efficiency.