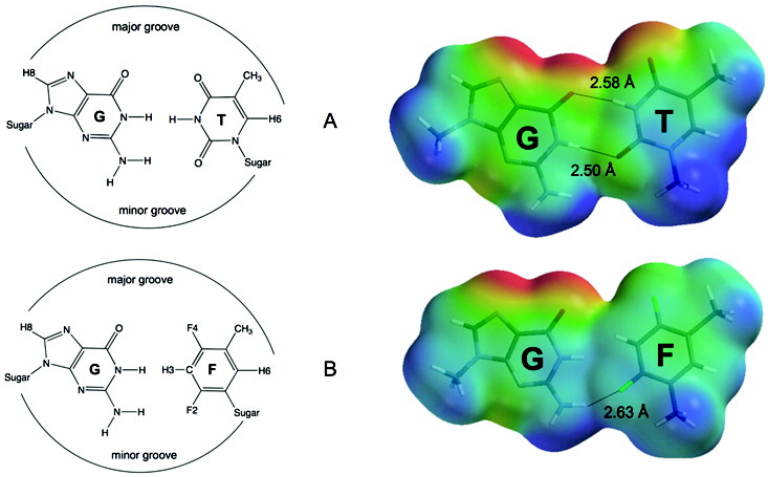
Article

Solution Structure of a DNA Duplex Containing a Guanine–Difluorotoluene Pair: A Wobble Pair without Hydrogen Bonding?

Danielle A. Pfaff, Kristine M. Clarke, Timothy A. Parr, Joanna M. Cole, Bernhard H. Geierstanger, Deborah C. Tahmassebi, and Tammy J. Dwyer

J. Am. Chem. Soc., **2008**, 130 (14), 4869-4878 • DOI: 10.1021/ja7103608



More About This Article

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

- Supporting Information
- Links to the 3 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 03/15/2008

Solution Structure of a DNA Duplex Containing a Guanine−Difluorotoluene Pair: A Wobble Pair without Hydrogen Bonding?

Danielle A. Pfaff,[†] Kristine M. Clarke,[†] Timothy A. Parr,[†] Joanna M. Cole,[†] Bernhard H. Geierstanger,[‡] Deborah C. Tahmassebi,*[,†] and Tammy J. Dwyer*[,†]

Department of Chemistry and Biochemistry, University of San Diego, San Diego, California 92110, and Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, California 92121

Received November 15, 2007; E-mail: tdwyer@sandiego.edu

Abstract: The incorporation of synthetic nucleoside analogues into DNA duplexes provides a unique opportunity to probe both structure and function of nucleic acids. We used ^{1}H and ^{19}F NMR and molecular dynamics calculations to determine the solution structures of two similar DNA decamer duplexes, one containing a central G-T mismatched or "wobble" base pair, and one in which the thymine in this base pair is replaced by difluorotoluene (a thymine isostere) creating a G-F pair. Here, we show that the non-hydrogen-bonding G-F pair stacks relatively well into the helix and that the distortions caused by each non-Watson−Crick G-T or G-F base pair are quite localized to a three base pair site around the mismatch. A detailed structural analysis reveals that the absence of hydrogen bonding introduces more dynamic motion into the G-F pair relative to G-T and permits the G-F pair to exhibit stacking and conformational features characteristic of both a Watson−Crick base pair (on the guanine containing strand) and a wobble base pair (on the strand containing the difluorotoluene). We used these results to posit a rationale for recognition and repair of mismatch sites in DNA.

Introduction

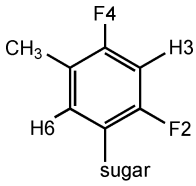
Double-stranded DNA is a robust biopolymer that can accommodate a host of perturbations away from standard Watson−Crick base pairing. Efforts to understand the range of factors responsible for DNA structure, stability, and function have exploited this tolerance while systematically deconstructing the very rules for double-helix formation put forth by Watson and Crick.[1,2] Successful approaches have centered around altering hydrogen-bonding patterns,[3−7] eliminating hydrogen-bonding interactions entirely,[8−13] and mediating base pair formation with metal cations.[14−16] Importantly, replacing natural nucleobases with aromatic molecules designed to mimic the steric and electronic properties of these DNA building blocks has helped tease out the details of DNA replication[17−25] and repair.[26−31] Solution structures of several DNA duplexes containing non-polar isosteres paired with the natural comple-

[†] University of San Diego.
[‡] Genomics Institute of the Novartis Research Foundation.

(1) Watson, J. D.; Crick, F. H. C. *Nature (London, U.K.)* **1953**, *171*, 737−738.
(2) Watson, J. D.; Crick, F. H. C. *Nature (London, U.K.)* **1953**, *171*, 964−967.
(3) Hoops, G. C.; Zhang, P.; Johnson, W. T.; Paul, N.; Bergstrom, D. E.; Davisson, V. J. *Nucleic Acids Res.* **1997**, *25*, 4866−4871.
(4) Benner, S. A. *Acc. Chem. Res.* **2004**, *37*, 784−797.
(5) Sismour, A. M.; Benner, S. A. *Nucleic Acids Res.* **2005**, *33*, 5640−5646.
(6) Yang, Z.; Hutter, D.; Sheng, P.; Sismour, A. M.; Benner, S. A. *Nucleic Acids Res.* **2006**, *34*, 6095−6101.
(7) Yang, Z.; Sismour, A. M.; Sheng, P.; Puskar, N. L.; Benner, S. A. *Nucleic Acids Res.* **2007**, *35*, 4238−4249.
(8) Schweitzer, B. A.; Kool, E. T. *J. Org. Chem.* **1994**, *59*, 7238−7242.
(9) Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863−1872.
(10) Schweitzer, B. A.; Sheils, C. J.; Ren, X.-F.; Chaudhuri, N. C.; Kool, E. T. Hydrophobic Nucleoside Isosteres as Biophysical Probes of Noncovalent Interactions: Analysis of Structure and Electrostatics. In *Biological Structure and Dynamics*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: Albany, NY, 1996; Vol. 2, pp 209−216.
(11) Matray, T. J.; Kool, E. T. *J. Am. Chem. Soc.* **1998**, *120*, 6191−6192.
(12) Smirnov, S.; Matray, T. J.; Kool, E. T.; de los Santos, C. *Nucleic Acids Res.* **2002**, *30*, 5561−5569.
(13) Nakano, S.; Uotani, Y.; Uenishi, K.; Fujii, M.; Sugimoto, N. *Nucleic Acids Res.* **2005**, *33*, 7111−7119.
(14) Atwell, S.; Meggers, E.; Spraggon, G.; Schultz, P. G. *J. Am. Chem. Soc.* **2001**, *123*, 12364−12367.
(15) Zimmerman, N.; Meggers, E.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 13684−13685.
(16) Shionoya, M.; Tanaka, K. *Curr. Opin. Chem. Biol.* **2004**, *8*, 592−597.
(17) Horlacher, J.; Hottinger, M.; Podust, V. N.; Hubscher, U.; Benner, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329−6333.
(18) (a) Guckian, K. M.; Kool, E. T. *Angew. Chem., Int. Ed.* **1997**, *109*, 2942−2945. (b) Ibid, *Angew. Chem., Int. Ed.* **1998**, *36*, 2825−2828.
(19) Moran, S.; Ren, R. X.-F.; Rumney, S.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056−2057.
(20) Morales, J. C.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 950−954.
(21) Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angew. Chem., Int. Ed.* **2000**, *39*, 990−1009.
(22) Kool, E. T. *Ann. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 1−22.
(23) Zhang, X.; Lee, I.; Berdis, A. J. *Org. Biomol. Chem.* **2004**, *2*, 1703−1711.
(24) Kirby, T. W.; DeRose, E. F.; Beard, W. A.; Wilson, S. H.; London, R. E. *Biochemistry* **2005**, *44*, 15230−15237.
(25) Leconte, A. M.; Matsuda, S.; Hwang, G. T.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2006**, *45*, 4326−4329.
(26) Fox, K. R.; Allinson, S. L.; Sahagun-Krause, H.; Brown, T. *Nucleic Acids Res.* **2000**, *28*, 2535−2540.
(27) Drotschmann, K.; Yang, W.; Brownewell, F. E.; Kool, E. T.; Kunkel, T. A. *J. Biol. Chem.* **2001**, *276*, 46225−46229.
(28) Schofield, M. J.; Brownewell, F. E.; Nayak, S.; Du, C.; Kool, E. T.; Hsieh, P. *J. Biol. Chem.* **2001**, *276*, 45505−45508.
(29) Vallur, A. C.; Feller, J. A.; Abner, C. W.; Tran, R. K.; Bloom, L. R. *J. Biol. Chem.* **2002**, *277*, 31673−31678.
(30) Francis, A. W.; Helquist, S. A.; Kool, E. T.; David, S. S. *J. Am. Chem. Soc.* **2003**, *125*, 16235−16242.
(31) Begley, T. J.; Haas, B. J.; Morales, J. C.; Kool, E. T.; Cunningham, R. P. *DNA Repair* **2003**, *2*, 107−120.

ment have been determined by NMR spectroscopy[32,33] and suggest that, while the absence of hydrogen bonding may destabilize the helix energetically, the global DNA structure remains intact and displays fairly localized distortions at the substitution site. The non-hydrogen-bonded pairs show stacking patterns and backbone conformations not too different from their natural counterparts. The incorporation of a wholly non-natural pair into a DNA duplex[34] likewise produces a dynamic situation, whereby the local structure is dictated by multiple conformations of the aromatic components.

Non-standard DNA base pairings can occur through a variety of mechanisms such as recombination, synthesis errors, deamination, and methylation. The processes by which mismatched pairs are recognized and corrected by repair systems have long been of interest.[35−38] While no single mechanism can be responsible for the diversity of events associated with DNA recognition and repair, most agree that structural changes as subtle as breathing, symmetry, and/or conformational dynamics may be critical factors.[37,39−44] Given the structural similarities displayed by DNA pairs formed between a natural nucleoside and a non-polar, non-hydrogen-bonding isostere with the Watson−Crick base pair from which it derives, we were curious to know whether a non-hydrogen-bonding pair derived from a non-Watson−Crick base pair (i.e., a "wobble" pair) would retain the structural features typical of such a mismatch.



1, "F"

We used ^{1}H and ^{19}F NMR and molecular dynamics (MD) calculations to determine the solution structures of two DNA duplexes of nearly identical sequence. One duplex contains a single G-T mismatch, while the other duplex contains a difluorotoluene (**1**) moiety to replace the G-T wobble pair with a G-F pair. These duplexes are henceforth referred to as (GT) and (GF).

5'-CCAAGCTTCC-3'	5'-CCAAGCTTCC-3'
3'-GGTTTGAAGG-5'	3'-GGTTFGAAGG-5'
GT	GF

Difluorotoluene is a well-studied[45−47] steric mimic of thymine that does not form hydrogen bonds with adenine (or any complementary base) yet contributes to duplex stability via stacking interactions. Indeed, an A-F pair confers relatively little structural change to canonical B-DNA nor does it disrupt DNA polymerase activity.[32] We detail here the structural similarities and differences between GT and GF and provide evidence supporting the claim that, at least in the case of mismatch correction, the repair system is capable of sensing esoteric structural features of the DNA substrate leading to mispair recognition.

Materials and Methods

Sample Preparation. The oligomers d(CCAAGCTTCC), d(GGAAGTTTGG), and d(GGAAGFTTGG) were synthesized and HPLC purified by Isis Pharmaceuticals, Inc. Oligomer concentrations were determined from UV absorbance at 260 nm, in which the extinction coefficient for difluorotoluene was estimated to be the same as thymine. The oligomers were mixed in a 1:1 stoichiometry, annealed at 85 °C, and then dialyzed against 2 L of 1.0 M NaCl, 2 L of 0.50 M NaCl, and 2 L of deionized water in a dialysis apparatus using a membrane with a 1 kDa molecular weight cutoff. The purity of the duplex DNA sample was greater than 95% based on NMR.

NMR samples were prepared by dissolving a particular DNA duplex in 600 μL of 10 mM sodium phosphate buffer (pH 7.0) containing 100 mM NaCl and 0.1 mM EDTA and then lyophilizing and redissolving in 99.99% D_2O (Cambridge Isotope Laboratories) to a final volume of 600 μL. The total DNA strand concentration in the resulting samples was 2 mM. We used the following numbering scheme to describe the duplexes in these studies:

```
5'- C1  C2  A3  A4   G5   C6  T7  T8  C9  C10 -3'
3'- G20 G19 T18 T17 T/F16 G15 A14 A13 G12 G11 -5'
```

Thermal Denaturation Studies. Melting studies were performed on three DNA duplexes (GC, GT, and GF) at a 5 μM strand concentration in 1 mL of buffer containing 20 mM sodium phosphate and 10 mM NaCl at pH 7.0. The increase in absorbance at 260 nm was monitored on a Uvikon XL UV−vis spectrophotometer as samples were heated at a rate of 0.5°/min. Melting temperatures were determined from fits to individual melting curves.

NMR Spectroscopy. Proton NMR spectra in D_2O for each duplex sample were acquired on a Varian Inova 500 MHz spectrometer. NOESY, DQF-COSY, and TOCSY spectra were acquired using the TPPI method of phase cycling.[48] Data were collected at 15 °C (GF) and 20 °C (GT) to resolve as much cross-peak overlap as possible in the spectra. For signal assignments in each duplex, NOESY spectra with a mixing time of 300 ms were collected with a spectral width of 5913 Hz and 2048 complex points in t_2 and 512 t_1 increments (zero filled to 2048 on processing). For each t_1 value, 64 scans were averaged using a recycle delay of 2 s. Presaturation was applied during the recycle delay and the mixing time to suppress the residual HOD resonance. DQF-COSY spectra were collected with 2048 complex points in t_2 and 512 t_1 increments (zero filled to 2048 points on processing) with a spectral width of 5913 Hz. For each t_1 value, 48−80 scans were collected using a recycle delay of 2 s with presaturation of the HOD resonance. TOCSY spectra were collected with 1024 complex points

(32) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 954−958.
(33) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 6841−6847.
(34) Matsuda, S.; Fillo, J. D.; Henry, A. A.; Rai, P.; Wilkens, S. J.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Schultz, P. G.; Spraggon, G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2007**, *129*, 10466−10473.
(35) Modrich, P. *Ann. Rev. Biochem.* **1987**, *56*, 435−466.
(36) Salsbury, F. R.; Clodfelter, J. E.; Gentry, M. B.; Hollis, T.; Scarpinato, K. D. *Nucleic Acids Res.* **2006**, *34*, 2173−2185.
(37) Iyer, R. R.; Pluciennik, A.; Burdett, V.; Modrich, P. L. *Chem. Rev.* **2006**, *106*, 302−323 and references therein.
(38) Parker, J. B.; Bianchet, M. A.; Krosky, D. J.; Friedman, J. I.; Amzel, L. M.; Stivers, J. T. *Nature (London, U.K.)* **2007**, *449*, 433−438.
(39) Brown, T.; Hunter, W. N.; Kneale, G.; Kennard, O. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 2402−2406.
(40) Hunter, W. N.; Brown, T.; Anand, N. N.; Kennard, O. *Nature (London, U.K.)* **1986**, *320*, 552−555.
(41) Hunter, W. N.; Brown, T.; Kennard, O. *Nucleic Acids Res.* **1987**, *15*, 6589−6606.
(42) Allawi, H. T.; SantaLucia, J., Jr. *Nucleic Acids Res.* **1998**, *26*, 4925−4934.
(43) Bhattacharya, P. K.; Cha, J.; Barton, J. K. *Nucleic Acids Res.* **2002**, *30*, 4740−4750.
(44) Spielmann, H. P. *J. Am. Chem. Soc.* **2004**, *126*, 583−590.
(45) Kool, E. T.; Sintim, H. O. *Chem. Commun. (Cambridge, U.K.)* **2006**, 3665−3675 and references therein.
(46) Wang, X.; Houk, K. N. *Chem. Commun. (Cambridge, U.K.)* **1998**, 2631−2632.
(47) Cubero, E.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2000**, *122*, 6891−6899.
(48) Drobny, G.; Pines, A.; Sinton, S.; Weitekamp, D. P.; Wemmer, D. E. *Faraday Div. Chem. Soc. Symp.* **1979**, *13*, 49−174.

in t_2 and 512 t_1 experiments (zero filled to 1024 points on processing) with a spectral width of 4504 Hz and mixing times of 50−75 ms. For each t_1 value, 64 scans were collected using a recycle delay of 2 s with presaturation of the HOD resonance. All spectra were transferred to a PC laptop and processed with Felix (Accelrys).

For generating distance constraints, NOESY spectra in D_2O were acquired at 15 °C (GF) or 20 °C (GT) using TPPI on a Varian Inova 500 MHz spectrometer. The spectra were collected with mixing times of 50, 100, and 200 ms with 2048 complex points in t_2 and 512 t_1 experiments (zero filled to 2048 on processing) and a spectral width of 5913 Hz. For each t_1 value, 64 scans were signal averaged using a recycle delay of 4 s with presaturation of the HOD resonance. The 2-D data were apodized with a skewed sine bell function in both dimensions (800 points, phase 60°, skew 0.5−0.7 in t_2; 512 points, phase 60°, skew 0.5−0.7 in t_1). The first row of the data matrix was multiplied by a factor of 0.5 prior to Fourier transformation in t_1 to suppress t_1 ridges. NOE cross-peaks were then integrated manually with Felix on a PC laptop.

Additional NOESY spectra in H_2O were acquired for both duplexes at 283 K on an Avance 600 MHz instrument equipped with a $^1H/^{13}C/^{15}N$−-TXI cryoprobe (Bruker Biospin) using 48 transients in each of the 512 t_1 experiments with a spectral width of 22 ppm, a mixing time of 200 ms, a recycle delay of 2 s, and excitation sculpting with gradients for water suppression.[49] Similarly, a 200 ms mixing time NOESY in H_2O for the GF duplex was also recorded at 277 K but on an Avance 400 MHz instrument equipped with a $^1H/^{13}C/^{19}F/^{31}P$-QNP cryoprobe (Bruker Biospin). Water suppression was accomplished using excitation sculpting with gradients.[49] These NOESY spectra were used for comparison with a $^1H-^{19}F$ NOESY (277 K, 512 scans for each of the 50 t_1 experiments, 22 ppm spectral width for proton and 12 ppm for fluorine dimension, 500 ms mixing time) recorded on the same instrument as described previously.[50] A 10 Hz line broadening function was applied in the first dimension. ^{19}F 1-D NMR spectra were acquired with 512 scans, 12 ppm spectral width, 8192 points, and proton decoupling and were processed with exponential multiplication with 3 Hz line broadening. Exponential multiplication with 1 Hz line broadening was applied to ^1H 1-D NMR spectra recorded using excitation sculpting with gradients.[49] Fluorine chemical shifts were determined relative to external 0.05% trifluorotoluene in $CDCl_3$ at −62.74.

Structure Calculations. To generate distance constraints, the off-diagonal NOE cross-peaks of each of the three assigned NOESY spectra were integrated using Felix, creating a total of three peak intensity sets for each duplex. The peak volumes were converted to distances using the cytosine H5−H6 distance of 2.46 Å as a reference. Depending on the trend in distances as a function of mixing time, distances were classified as very strong (1.8−2.2 Å), strong (2.2−2.8 Å), medium (2.8−4.0 Å), weak (4.0−4.5 Å), or very weak (4.5−5.0 Å). All lower bounds for distance restraints were set at 1.8 Å, and upper bounds were set at the top of the category range. Dihedral angles (torsions) were loosely restrained based upon close inspection of the DQF-COSY spectra.[51−53] In the C1′H to C2′H/C2″H region (approximately 5.0−6.4 ppm in F1 and 1.8−2.9 ppm in F2), if the distance between outer peaks in the F1 dimension was greater than 14 Hz, the δ torsion angle was restrained between 110 and 170°. For the GT duplex, a total of 347 constraints was applied (including Watson−Crick hydrogen-bonding constraints, 222 NMR-derived distance restraints, and torsion restraints for each sugar moiety). For the GF duplex, a total of 357 constraints was applied (including Watson−Crick hydrogen bonding (excepting the G-F pair), 239 NMR-derived distance restraints, and torsion restraints for each sugar moiety except F16). In the GT duplex and for all bases except difluorotoluene in the GF duplex, each imino N−H proton was observed in NOESY spectra acquired in H_2O and displayed broadening behavior characteristic of Watson−Crick hydrogen-bonded base pairing upon increasing the temperature. This provided justification for including hydrogen-bond restraints in the MD simulations. The lack of hydrogen bonding in the G-F pair (vide infra) precluded using such restraints.

Table 1. Melting Temperatures Derived from Thermal Denaturation Curves[a]

DNA Duplex	Designation	Tm (°C)
5′-CCAAGCTTCC-3′ 3′-GGTTCGAAGG-5′	GC	53.4
5′-CCAAGCTTCC-3′ 3′-GGTTTGAAGG-5′	GT	38.9
5′-CCAAGCTTCC-3′ 3′-GGTTFGCCGG-5′	GF	30.2

[a] Duplex concentrations were 5 μM in 1 M NaCl, 10 mM sodium phosphate buffer (pH 7.0), and 0.1 mM EDTA.

The approach for computing structures used in this study was patterned after that of Smith et al.[54] Force field parameters for difluorotoluene were calculated using Gaussian 98.[55] All energy minimization and restrained molecular dynamics (rMD) calculations were performed with the SANDER module of AMBER 9.[56] The NAB molecular manipulation language[57] was used to create a series of 40 starting DNA structures for each duplex. These structures differed in the four helical parameters inclination, rise, twist, and *x*-displacement to give a diverse set of A- and B-form DNA molecules. For each duplex, every member of the ensemble was subjected to 1000 steps of steepest descent energy minimization followed by slow equilibration to 300 K while applying harmonic constraints with a force constant of 10.0 kcal mol^{-1} Å^2. The structures were further equilibrated at 300 K as harmonic restraints were reduced to 1 kcal mol^{-1} Å^2. The resulting DNA structures for each duplex had a mean pairwise root-mean-square deviation (rmsd) of 4.79 Å. The fully equilibrated structures were then subjected to two rounds of restrained MD. In the first cycle, the temperature was held constant at 300 K, while constraints were increased over 3 ps to full strength, where they remained for an additional 17 ps of rMD. The refinement was completed with a cycle of rMD simulated annealing using all 347 constraints (GT) or 357 constraints (GF). In each case, the duplex was heated to 700 K over 2 ps and held at this temperature for 3 ps of dynamics. The system was then cooled to 0 K over 15 ps. In each 20 ps simulation, the force constant for all distance constraints was increased linearly from 0 to 100 kcal mol^{-1} Å^2 over 3 ps, remaining at 100 kcal mol^{-1} Å^2 for the final 17 ps of the simulation, while the force constants for Watson−Crick hydrogen bonding and torsions were held at 32.0 kcal mol^{-1} Å^2 during each 20 ps simulation. All structures were then subjected to 100 steps of steepest descent energy minimization. Helical parameters for the final structures were derived using CURVES 5.1.[58]

Results

Thermal Denaturation Studies. Table 1 summarizes the melting temperatures of the GT and GF duplexes discussed in this study as well as the control GC duplex d(CCAAGCTTCC)·

(49) Hwang, T. L.; Shaka, A. J. *J. Magn. Reson, Ser. A* **1995**, *112*, 275−279.
(50) Scott, L. G.; Geierstanger, B. H.; Williamson, J. R.; Hennig, M. *J. Am. Chem. Soc.* **2004**, *126*, 11776−11777.
(51) Baleja, J. D.; Pon, R. T.; Sykes, B. D. *Biochemistry* **1990**, *29*, 4828−4839.
(52) Gronenborn, A. M.; Clore, G. M. *Biochemistry* **1989**, *28*, 5978−5984.
(53) Clore, G. M.; Oschkinat, H.; McLaughlin, L. W.; Benesler, F.; Happ, C. S.; Happ, E.; Gronenborn, A. M. *Biochemistry* **1988**, *27*, 4185−4197.
(54) Smith, J. A.; Gomez-Paloma, L.; Case, D. A.; Chazin, W. J. *Magn. Reson. Chem.* **1996**, *34*, 147−155.
(55) Pople, J. A. et al. *Gaussian 98*, revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.
(56) Case, D. A. et al. *AMBER 9*; University of California, San Francisco: San Francisco, 2006.
(57) Macke, T. Nab, a language for molecular manipulation. Ph.D. Thesis, Scripps Research Institute, La Jolla, CA, 1996.
(58) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63−91.

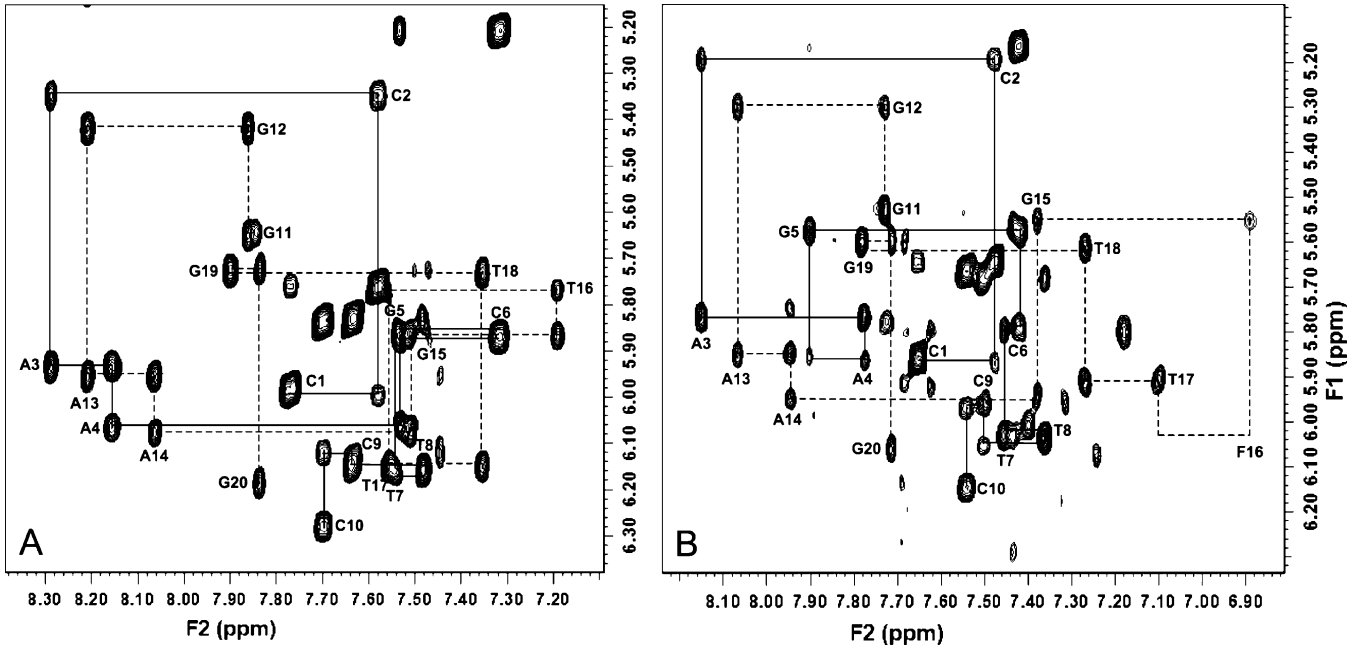


Figure 1. Deoxyribose H1′ to aromatic portion of the NOESY spectra of GT (A) at 20 °C and GF (B) at 15 °C, both at a mixing time of 300 ms. The sequential connectivity for H1′ to H6/H8 for residues 1−10 is indicated by a solid line, while that for residues 11−20 is indicated by a dashed line.

(GGAACGTTGG) composed of all paired, natural nucleobases. The decreasing stability of the duplexes in the order of GC > GT > GF is not surprising given the loss of hydrogen bonding; however, the magnitude of reduction is greater for GC to GT than for GT to GF, perhaps correlating with a larger structural disruption in this order.

Signal Assignments and NMR Observations. Signals of exchangeable and non-exchangable protons of the GT and GF duplexes (with the exception of deoxyribose C5′H and C5″H) were assigned using standard procedures.[59,60] The assignments and NMR data are summarized in the Supporting Information.

All imino protons were visible for each duplex at 5 °C; however, the peaks corresponding to G11 and G15 NH were broadened at this temperature. In addition, G5 NH and T17 NH signals in the GF duplex were also quite broad at 5 °C and barely visible at 10 °C. Assignments were confirmed using NOESY spectra acquired in H_2O (Supporting Information).

The sequential assignments of the aromatic base protons and deoxyribose C1′H for the GT duplex are shown in Figure 1A at 20 °C. The upfield shift for H6 of T16 involved in the mispair distinguishes it from the other thymines and was useful in assigning the T16 methyl protons. Figure 1B shows the same region for the GF duplex. The NOESY and DQF-COSY spectra support one predominant form of the GF duplex in solution. The aromatic H6 of F16 (see **1**) is shifted farther upfield than H6 of T16 in GT and displays the typical sequential NOE connectivity to the nearby C1′H of F16 and G15; albeit, the peaks are a bit weaker for this residue than other bases. We interpret this as evidence of increased dynamics in the region of F16 and a requisite broadening of its resonances. The methyl protons for T7, T8, T17, and F16 were unambiguously assigned in a manner similar to that of GT. Sequential connectivity of aromatic protons with C2′H and C2″H confirmed the analysis of the C1′H-C2′H and C1′H-C2″H cross-peak patterns in the DQF-COSY and TOCSY spectra to unambiguously assign these protons. The ^{19}F spectrum of GF shows only two peaks, in a 1:1 ratio. Heteronuclear $^{1}H-^{19}F$ 2-D NOESY spectra (Figure 2) of GF permitted unambiguous assignment of these signals as F16 F2 (−120.1 ppm) and F4 (−113.6 ppm). The resonance for F16 F2 is broader than that for F4 and continues to broaden slightly with decreasing temperature. Over the temperature range of 4−27 °C, both the line width and chemical shift of F2 change more significantly than for F4. This observation is consistent with dynamics in the region of the difluorotoluene, possibly indicating that the area around F4 is more anchored, while the area around F2 is more fluxional. In other words, the major groove side of F16 may act as a hinge about which the difluorotoluene moiety pitches and/or sways.

Quality of the Structures. The family of structures used to represent each duplex was determined using the statistical analysis suggested by Smith et al.[53] For each duplex, the 40 structures resulting from rMD were placed in random order, and the mean all-atom pairwise rmsd was taken for the first two structures in the group, then the first three structures in the group, etc. This procedure was repeated 500 times, each time with a different random ordering of the 40 structures. The analysis predicts the minimum number of structures sufficient to represent the conformational space consistent with the experimental data, in the limit of the change in the standard deviation of rmsd with the addition of a structure to the family. It was determined that an ensemble of 10 structures was adequate to describe the GT duplex, while 15 structures were necessary to describe the GF duplex. The 10 structures (GT) and 15 structures (GF) in the final ensembles were chosen to minimize both the molecular mechanics (AMBER) energy and the constraint violation energy.

A summary of the energy and rmsd characteristics of the ensembles of structures for the GT and GF duplexes is presented in Table 2. Given that the 40 equilibrated starting structures represented a range of A-DNA and B-DNA conformations and

(59) Hare, D. R.; Wemmer, D. E.; Chou, S.-H.; Drobny, G.; Reid, B. R. *J. Mol. Biol.* **1983**, *171*, 319−336.
(60) Wuthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.

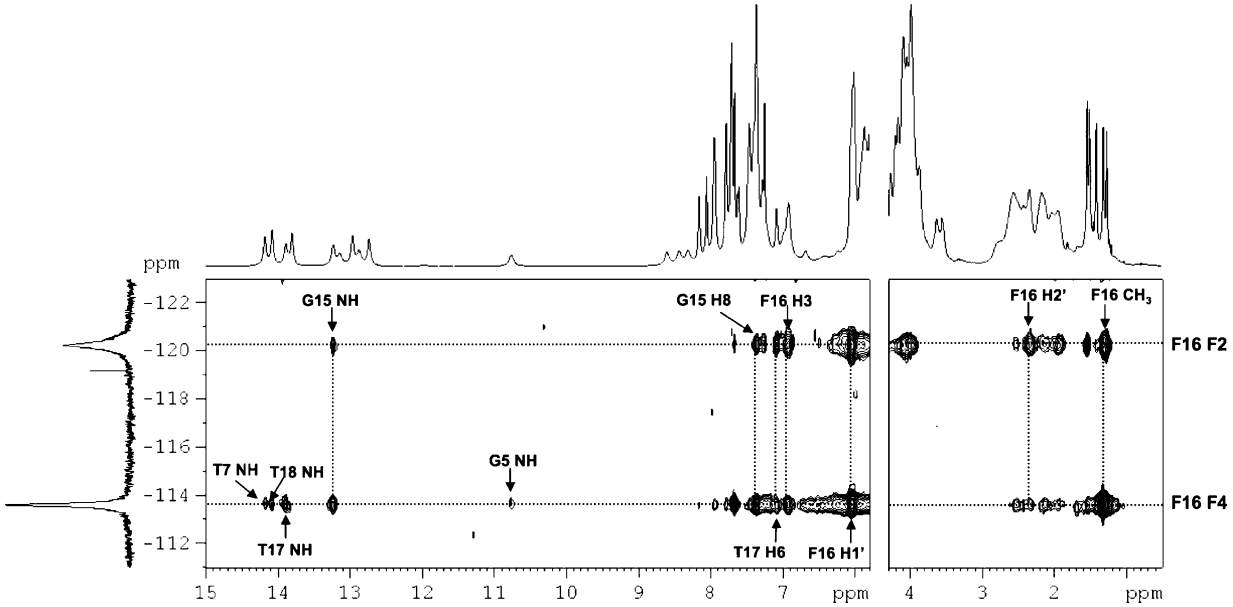


Figure 2. Heteronuclear $^{1}H-^{19}F$ NOESY spectrum of GF at 4 °C (acquisition parameters described in Materials and Methods). The chemical shifts of F2 and F4 of the difluorotoluene are −120.1 and −113.6 ppm, respectively.

Table 2. Summary of Energies, rmsd Values, and Violations for Ensembles of Structures

	GT	GF
Molecular mechanics energies (kcal)		
E_{Amber}	−4340.4 ± 2.7	−4262.9 ± 2.6
E_{Viol}	12.9 ± 0.6	24.7 ± 0.5
Average pairwise rmsd (Å)		
DNA	1.45	1.33
central six base pairs	0.970	0.93
Distance violations (Å)		
$0.05 < d \leq 0.10$	18	31
$0.10 < d \leq 0.20$	3	5
$0.20 \leq d$	0	0

helical parameters with an rmsd of 4.79 Å, the data in Table 2 show strong structural convergence for GT and GF with all-atom, pairwise rmsd values of 1.45 Å (rms difference from the mean structure of 0.97 Å) and 1.33 Å (rms difference from the mean structure of 0.91), respectively. As is generally the case, removing the last two base pairs from either end of the GT or GF duplex resulted in a marked decrease in the rmsd for each structure. For the central six base pairs, the rmsd for the GT ensemble was 0.97 Å (rms difference from the mean structure of 0.62 Å), and the central six base pairs of the GF ensemble had an rmsd of 0.93 Å (rms difference from the mean structure of 0.63 Å). The superposition of the family of 10 structures describing the GT duplex is shown in Figure 3A, while the 15 structures characterizing the GF duplex are shown in Figure 3B. The final ensembles of structures for GT and GF have total restraint violations summing 12.9 ± 0.6 and 24.7 ± 0.5 kcal, respectively. These violations amount to just 0.3% (GT) and 0.6% (GF) of the total energy of the systems.

Structural Features of the Duplexes. (1) GT. The average solution structure describing the GT duplex is shown in Figure 3C. The stacking interactions in GT are further illuminated upon inspection of the electrostatic potential surfaces of the central three base pairs of the average structure as presented in Figure 4A. The coplanar stacking geometry of the base pairs is enforced by the close contacts of the van der Waals surfaces in addition

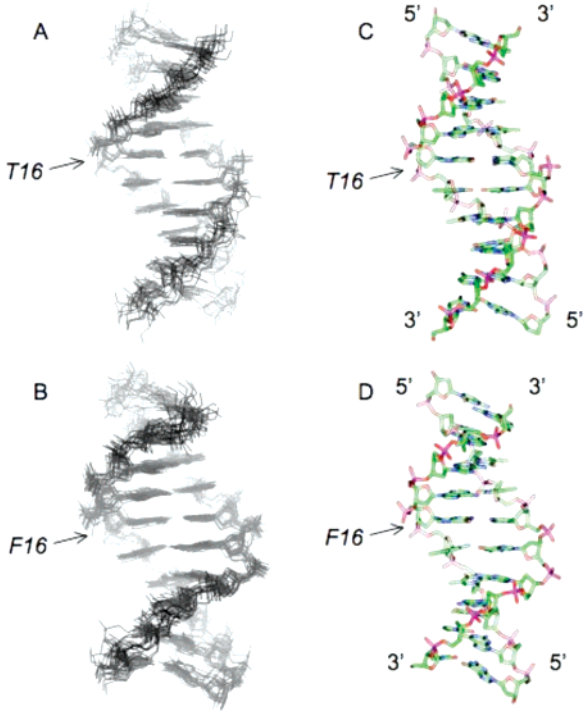


Figure 3. Superposition of family of 10 structures describing the GT duplex (A) obtained by rMD, superposition of the family of 15 structures describing the GF duplex (B) obtained by rMD, and a comparison of the average structures for GT (C) and GF (D).

to the electrostatic repulsion of these surfaces in the interbase region of the helix. The pyrimidine ring of T16 is coplanar with G15 and stacks onto the five-membered ring of the purine. T16 is also coplanar with T17; however, T16 O2 stacks onto the pyrimidine ring of T17, placing the ring of T16 directly over the T17 O4 and T17 methyl group. This is similar to the stacking geometry of the bases in the similar region of the crystal structure of a comparable duplex $d(CCAAGCTTGG)_2$ determined by Dickerson and colleagues (PDB entry 158D).[61] In 158D, residue 16 is a cytosine (which base pairs in Watson–Crick fashion with G5), and C16 O2 stacks onto the pyrimidine

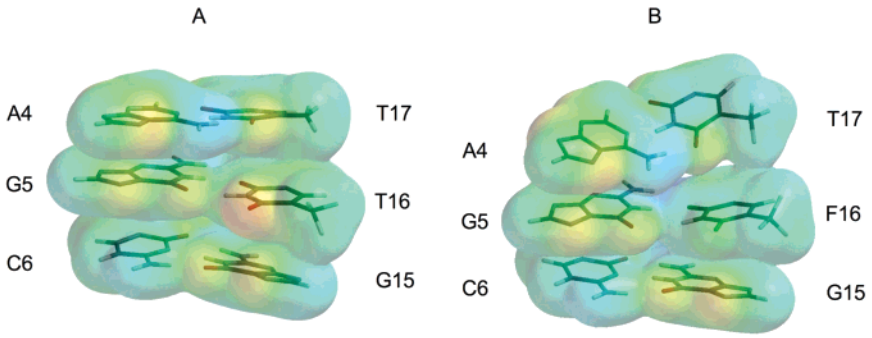


Figure 4. Central three base pair regions for the GT (A) and GF (B) average duplexes in which the electrostatic potentials have been mapped onto the electron density surfaces calculated at the PM3 semiempirical level. The same potential range (−75.0 to +80.0 from red to blue) was used for each structure.

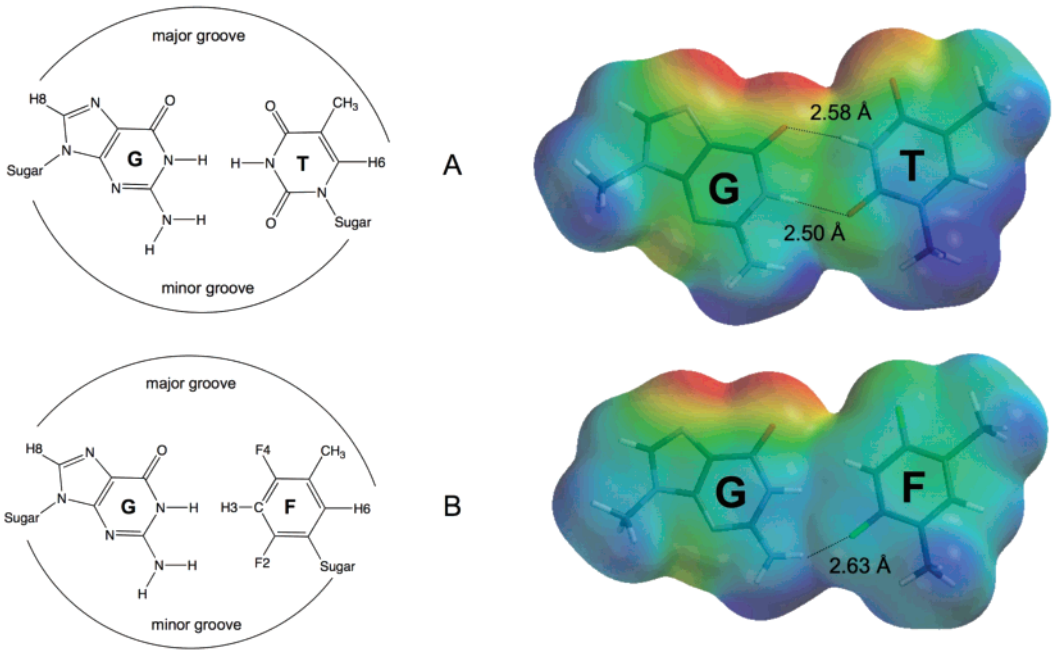


Figure 5. Schematic (left) and electrostatic potential surfaces (right) of the G-T base pair (A) and G-F pair (B) in the average duplex structures calculated using density functional theory (DFT) at the B3LYP/6-31G* level. The same potential range (−70.0 to +15.0 from red to blue) was used in each display. Hydrogen-bonding distances in the G-T wobble pair are also shown as well as the distance between F2 of F16 and the G5 amino group in the G-F pair. For simplicity, the deoxyribose (sugar) rings in each nucleotide moiety are replaced by methyl (R) groups.

ring of T17, placing the ring of C16 directly over T17 O4 and the T17 methyl group. The opposite strand in GT shows a quite different geometry. The pyrimidine ring of C6 is coplanar with G5 and stacks onto the six-membered ring of the purine. However, while G5 is coplanar with A4, it does not stack well, with G5 O6 skirting the edge of the A4 ring system. This is in contrast to 158D, in which the six-membered rings of A4, G5, and C6 are coplanar and stack onto each other.

The G5-T16 couple in this duplex exhibits wobble base pairing, whereby T16 is rotated into the major groove forming a hydrogen bond between G5 O6 and T16 N-H3 and also between G5 N-H1 and T16 O2 (see Figure 5A). An inspection of the electrostatic potential surface of the G-T base pair provides a visual representation of the hydrogen-bonding pattern. The negative electrostatic potential indicated by the green color in between G5 and T16 in Figure 5A denotes a sharing of electron density among hydrogen-bond donor and acceptor groups on the respective bases.

(2) GF. The NMR spectral data provide evidence that the difluorotoluene moiety is positioned inside the double helix, effectively forming a base pair with G5. Key NOEs in the heteronuclear $^{1}H-^{19}F$ NOESY (Figure 2) between both F2 and F4 of F16 and G15 and T17 imino protons (among others) and an NOE between G15 NH and F16 H3 (across the pair, data not shown) provide strong evidence of the intrahelix orientation of the difluorotoluene. The G5-F16 pair in this duplex adopts a somewhat different geometry than the G5-T16 base pair in GT. The G5-F16 pair is not precisely planar; rather, it experiences a higher propeller twist than the G-T base pair in GT. Figure 3D shows the average solution structure describing the GF duplex, while the stacking interactions and electrostatic potential surfaces of the central three base pairs of the average structure in GF are displayed in Figure 4B. In GF, G15 is coplanar with F16, and the six-membered ring of the purine stacks onto the phenyl ring of the difluorotoluene. In a manner similar to both 158D[61] and the GT duplex, F4 of F16 stacks directly over the pyrimidine ring of T17, and the phenyl ring of F16 stacks over T17 O4 and the T17 methyl group. On the opposite strand of

(61) Grzeskowiak, K.; Goodsell, D. S.; Kaczor-Grzeskowiak, M.; Cascio, D.; Dickerson, R. E. *Biochemistry* **1993**, *32*, 8923−8931.

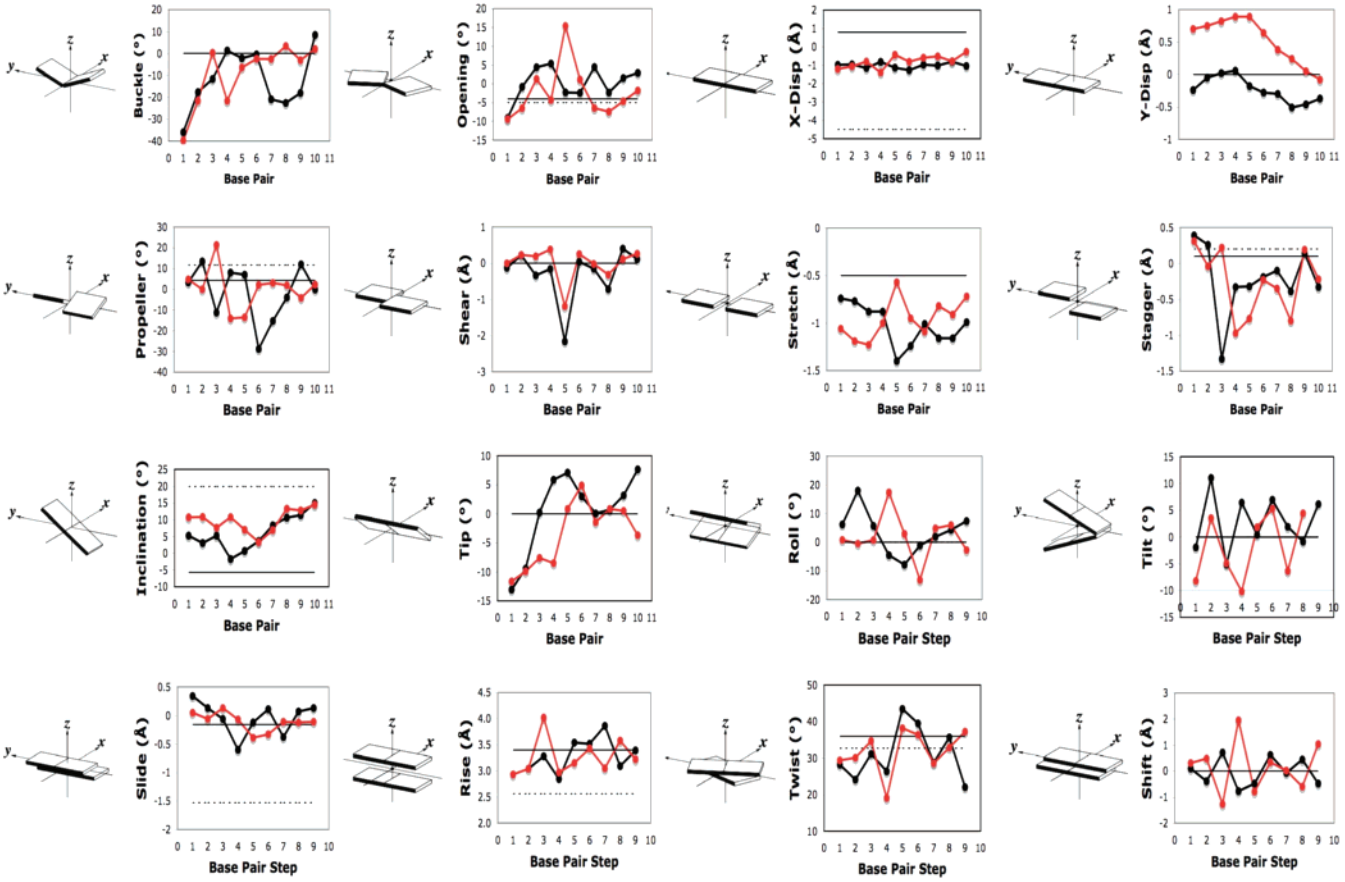


Figure 6. Helical parameters for the GT (black lines) and GF (red lines) average duplexes obtained using CURVES. Standard values for canonical B-DNA are indicated by solid horizontal lines, while those for A-DNA are indicated by dashed horizontal lines.

the GF duplex, the stacking pattern of A4-G5-C6 is identical to that in 158D,[61] whereby the six-membered ring of each base is consecutively stacked and fairly coplanar. We observe that F16 is not coplanar with T17 as the latter is substantially buckled toward G5-F16. An examination of the electrostatic potential surfaces of the central three base pairs of the GF duplex sheds light on the origin of this structural perturbation. In Figure 4B, it appears as though the regions of negative electrostatic potential associated with T17 O2 and O4 are oriented toward the region of high positive electrostatic potential between G5 and F16. The result is a closer van der Waals interaction than would be the case if A4-T17 were planar. The lack of hydrogen bonding and, hence, electron density in the central region of the helix precludes the electrostatic repulsion by neighboring base pairs. Instead, we postulate that a base pair adjacent to a non-natural base pair will be structurally disrupted in a sequence-dependent manner. The degree to which this effect will be manifested will depend upon the nature and position of the negative poles of polar groups on neighboring bases.

Although the mean distance between the putative hydrogen-bond donor (G5 NH-2) and acceptor (F16 F2) might suggest otherwise (see Figure 5B), the electrostatic potential surface implies that there is no hydrogen bonding between G5 and F16. Indeed, the region between G5 and F16 exhibits substantial positive character in the electrostatic potential. These results are synchronous with the findings of Wang and Houk,[46] who investigated the likelihood of hydrogen-bond formation in the adenine-difluorotoluene pair by computing electrostatic charges and analyzing electrostatic potential surfaces.

Helix Parameters. Both the GT and the GF duplexes in this study exhibited overall B-form DNA geometry with small deformations localized to the central three base pairs. The terminal residues of each strand have less reliable conformations due to a dearth of NOE data in these areas. The helical parameters that characterize the GT and GF solution structures are displayed in Figure 6. The large fluctuations in the buckle especially at the A4-T17 base pair of GF appear to result from the thymidine attempting to achieve better van der Waals contact with F16 and/or the G5-F16 pair itself. The negative propeller twist associated with A4-T17 and G5-F16 of GF is a consequence of the disruption in hydrogen bonding of the latter pair. The large negative propeller twist of the base pair on the 3′ side of the G-T wobble pair is a feature common to other structures of DNA duplexes containing these mismatches.[43] Note, however, that the GF duplex does not display this behavior in the C6-G15 base pair. Both duplexes show opening behavior similar to canonical B-DNA, with the exception of the G5-F16 pair of GF. The large positive value for opening in this pair correlates with the lack of hydrogen bonding. As well, the trend in the shear value across each duplex is quite regular except in the region of the G-T mismatch and G-F pair. The large negative values for shear correlate with the non-Watson Crick orientation of the bases in these pairs.

The deoxyribose ring conformations in the GT duplex have an average sugar pucker (pseudo-rotation phase angle) of 149° with no major interruptions with the exception of the fluxional terminal residues (see Figure 7A). By comparison, the GF duplex displays an average sugar pucker of 154° for the

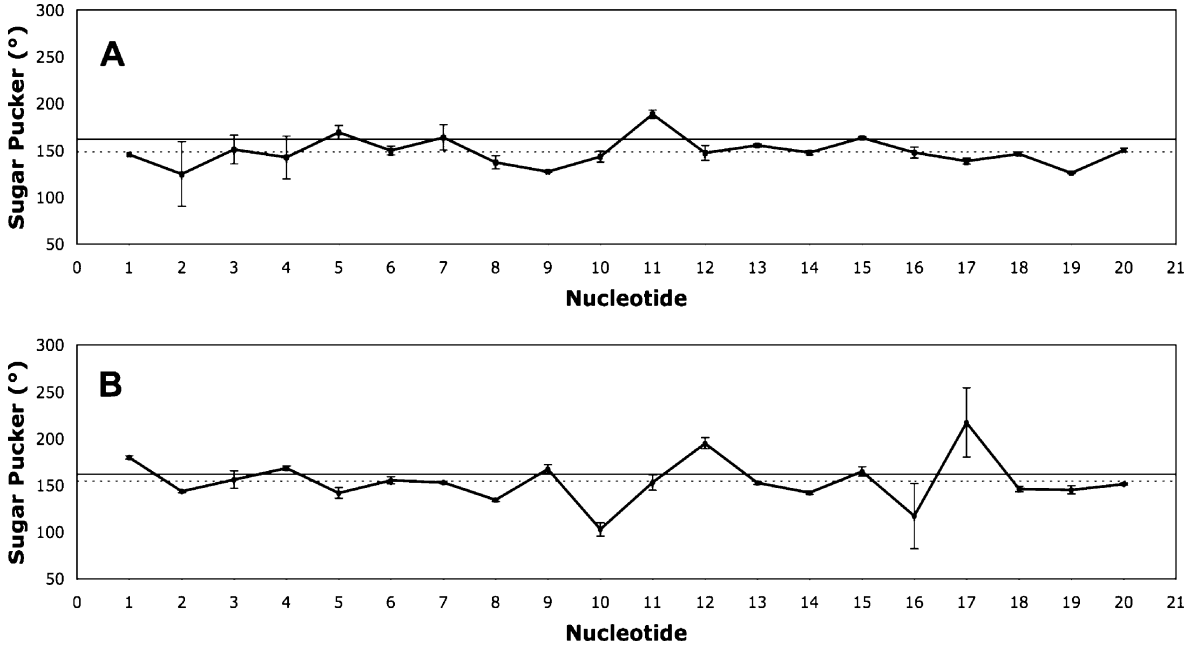


Figure 7. Comparison of the deoxyribose ring conformations in the GT (A) and GF (B) average duplexes. The solid horizontal line represents the expected value for B-form DNA, while that for A-form DNA is represented by a dashed horizontal line.

Table 3. Helical Parameters λ_1 and λ_2 in GT, GF, and $(CCAAGCTTGG)_2$[61] Duplexes

GT			GF			158D		
base pair	λ_1(deg)	λ_2(deg)	base pair	λ_1(deg)	λ_2(deg)	base pair	λ_1(deg)	λ_2(deg)
C1-G20	56.2 ± 1.1	56.2 ± 1.3	C1-G20	58.1 ± 1.0	54.7 ± 1.3	C1-G20	53.4	54.3
C2-G19	56.5 ± 0.8	54.8 ± 0.8	C2-G19	56.3 ± 0.8	51.7 ± 0.4	C2-G19	57.9	56.7
A3-T18	54.4 ± 1.1	58.7 ± 1.6	A3-T18	55.8 ± 1.1	52.7 ± 1.8	A3-T18	58.0	57.9
A4-T17	56.7 ± 0.7	57.9 ± 0.5	A4-T17	56.3 ± 1.3	51.9 ± 3.0	A4-T17	59.4	57.5
G5-T16	40.5 ± 0.6	66.4 ± 1.4	G5-F16	53.5 ± 2.2	70.0 ± 6.6	G5-C16	52.7	56.1
C6-G15	52.3 ± 1.2	51.9 ± 0.4	C6-G15	56.4 ± 0.4	53.5 ± 0.5	C6-G15	56.9	56.2
T7-A14	57.3 ± 0.9	57.0 ± 0.8	T7-A14	51.4 ± 0.7	51.4 ± 0.7	T7-A14	57.6	58.2
T8-A13	49.5 ± 1.2	58.5 ± 0.8	T8-A13	49.3 ± 0.7	52.8 ± 0.7	T8-A13	59.9	57.2
C9-G12	56.8 ± 0.9	53.1 ± 0.7	C9-G12	53.0 ± 0.9	51.7 ± 0.7	G9-C12	52.5	56.6
C10-G11	57.2 ± 1.6	55.9 ± 0.6	C10-G11	55.2 ± 1.2	52.0 ± 1.0	G10-C11	55.7	54.3

deoxyribose rings. The F16 and T17 residues show greater ranges in phase angle, most likely caused by an increased conformational flexibility around the non-hydrogen-bonded G5-F16 pair (see Figure 7B).

An analysis of the helical parameters λ_1 and λ_2 provides further evidence that all significant structural distortions in the GT and GF duplexes are localized to the mismatch sites. The angle formed by the purine N9 or pyrimidine N1 and C1′-C1′ intrabase pair vector defines λ. Each base pair may be characterized by two λ values, relating to each strand of the duplex. The measured values of λ_1 and λ_2 (the subscript 1 corresponds to residues 1−10 and 2 designates residues 11−20) are given in Table 3. Also provided in Table 3 are comparison values of λ_1 and λ_2 in 158D (d(CCAAGCTTGG)$_2$).[61] For B-DNA containing all Watson−Crick base pairs, a high degree of symmetry is observed in λ_1 and λ_2 values, approximately 55.9° each. Indeed, Table 3 shows that for the model duplex 158D, the range of λ values encompasses 52−59°. In B-DNA containing mismatched base pairs, the values of λ_1 and λ_2 tend to differ.[40−42] The data in Table 3 mirror this trend, with quite typical values for all residues except the G5-T16 base pair of GT and the G5-F16 pair in GF. Interestingly, while G5-T16 in GT shows λ values (40.5 and 66.4°) common in this type of mispair,[40−42] G5 of GF exhibits a λ value typical of a Watson−Crick base pair (53.5°), and F16 exhibits a λ value in the range expected for a wobble pair (70.0°).

Discussion

The solution structures of two DNA decamer duplexes (GT: d(CCAAGCTTCC)·(GGAAGTTTGG) and GF: d(CCAAGCTTCC)·(GGAAGFTTGG)), identical except for a single nucleotide, were determined with good precision by NMR spectroscopy and restrained MD (Figures 1−3). One duplex contains a single G-T mismatch that displays hydrogen bonding typical of a wobble base pair, while the other duplex replaces the thymidine in the mispair with its isostere mimic, difluorotoluene (F, **1**) (Figures 4 and 5). Measurements of the thermal stability of GT and GF as compared to the all Watson−Crick decamer duplex d(CCAAGCTTCC)·(GGAAGCTTGG) (Table 1) indicated that the G-F pair significantly destabilizes the duplex, even more than the G-T wobble pair. We were particularly interested to know whether the structural features that characterize the G-T wobble base pair would also be present in the G-F pair. The DNA double helix is remarkably accommodating of defects whether by nicks, gaps, base pair mismatches, covalent adduct formation, and myriad lesions. NMR studies[62] on a wide range of damaged DNAs have established that structural deformations occur largely in the vicinity of the defect and that the canonical

structure is restored within one base pair on either side. The structures obtained for the GT and GF duplexes in this work are consistent with these observations. As such, we will limit our discussion of the structural details in the present duplexes to the three base pair region centered on the G-T or G-F pair.

As compared to the G-F pair, the G-T base pair displays somewhat different stacking interactions with the adjacent base pairs (Figure 5). The static models generated in this study suggest that on average, the A4-T17/G5-T16/C6-G15 base pairs are fairly coplanar in orientation. However, with T16 thrust into the major groove to hydrogen bond with G5, A4-T17 can achieve a favorable stacking arrangement with only one of the adjacent base pairs, in this case C6-G15. The stacking of A4-T17/C6-G15 base pairs is quite poor in GT in contrast to the all Watson−Crick duplex 158D, d(CCAAGCTTGG)$_2$, studied by Dickerson et al.,[61] in which the A4-T17/G5-C16/C6-G15 base pairs are optimally stacked. Interestingly, the lack of hydrogen bonding in the G-F pair allows the mispaired bases on each strand to behave independently and facilitates more efficient stacking of G5 with both A4 and C6, similar to what is seen in 158D.[61] While the difluorotoluene stacks moderately well with G15, it stacks quite poorly with T17. In this way, the G-F pair exhibits structural features characteristic of both wobble and Watson−Crick base pairs. We postulate that since the more rigid wobble conformation of the G-T pair is not enforced by hydrogen bonding in the G-F pair, favorable stacking interactions with neighboring bases can drive G5 to a lower energy configuration.

Coupled with the loss of stacking between the A4-T17/G5-F16 base pairs in GF is a pronounced buckle of A4-T17 toward the G-F pair that is not apparent in the GT structure. Inspection of the electrostatic potential surfaces of the A4-T17/G5-C16/C6-G15 region of the GT and GF duplexes provides some insight into the origin of this affect. In GT, the natural hydrogen bonding between base pairs places moderate electron density in the central region of the double helix (diffuse green in Figure 5A). The tight association of electron clouds of consecutive base pairs leads to fairly planar stacking with maximized van der Waals' interactions. In GF, however, the lack of hydrogen bonding between the G-F pair creates a sink of positive potential in this region of the double helix (localized blue color in Figure 5B) that serves to attract the electron-rich carbonyl oxygens of the neighboring thymine for optimum van der Waals' interactions between the base pairs. This necessarily leads to T17 orienting toward G5-F16 while remaining hydrogen bonded to A4.

The recognition and repair of mismatched base pairs in DNA by myriad proteins has long been the subject of study. Of likely importance to the specificity of mispair recognition is the flexibility of DNA especially as manifested in the backbone parameters. Kennard[39−41] used X-ray crystallographic models to characterize the molecular structures of DNA duplexes containing CA, GA, and G-T mismatches. While these structures display minimal overall distortion from canonical B-DNA, there are local perturbations in stacking and backbone conformations near the mismatch sites. Most interesting, the authors report that the C1′ to C1′ distance of the base pairs remains fairly invariant throughout the duplex. However, the angles between the glycosidic bond (N1 or N9 to C1′) and the C1′ to C1′ vector of the base pair, referred to as λ, differ at the mismatch site. For each base pair, two λ angles are measured depending on which strand the N−C1′ bond lies. Further, λ_1 and λ_2 for Watson−Crick base pairs display symmetry with both values in a narrow range. However, X-ray crystallographic-[39−41] and NMR-derived[42] structures reveal that λ values at mismatched sites show quite different trends. For DNA containing a G-T mispair, λ_1 and λ_2 differ greatly with values near 40 and 70°. This asymmetry in the λ value is most pronounced for the G-T mispair, while the λ values for a CA mismatch range from 46 to 72°. In fact, the degree of asymmetry in λ follows G-T > CA > GA, which parallels the efficiency of mispair recognition and repair in a number of studies.[63−65] Kennard et al.[39−41] postulated that the asymmetry in λ may be a key structural element for the recognition of mismatches by repair enzymes.

The G-F pair in the GF duplex exhibits structural features that tend to characterize both a Watson−Crick pair and a wobble pair. Therefore, a G-F pair has dual properties that may have implications for DNA recognition and repair. In 158D, d(CCAAGCTTGG)$_2$, the all Watson−Crick base paired duplex, the G5-C16 pair shows stacking interactions with adjacent bases on each strand typical of canonical B-DNA.[61] Further, the λ angles in 158D for all base pairs including G5-C16 average 52−59° (see Table 3). As discussed previously, the G-T base pair in the GT duplex displays the stacking patterns and asymmetry in λ angles characteristic of this type of mismatch (40.5 and 66.4°). This asymmetry is lost in the G-F pair in the GF duplex, suggesting that in the absence of hydrogen bonding to confine G5 into a wobble orientation, it occupies a position in the strand that optimizes stacking with adjacent bases and symmetrically disposes its location within the duplex, as if it were part of a Watson−Crick pair. This behavior is manifest in λ for G5 of 53.5°. F16, on the other hand, stacks in a similar fashion to T16 in the G-T wobble pair and exhibits a λ angle of 70.0°, similarly indicative of a mispaired base.

Replacing the thymine base of a G-T wobble pair in a DNA duplex with a difluorotoluene molecule has no observable effect on the global structure of the oligomer. Small structural differences are apparent in the immediate vicinity of the mismatch G-T or G-F pairs, one of which may, interestingly, play a role in some types of DNA mismatch repair. The mismatch endonuclease Vsr repairs G-T mispairs in a sequence-dependent and strand-specific manner.[26] Vsr has been shown to nick double-stranded DNA on the 5′-side of the thymine (underlined) in the mispair contained in the T̲AG or T̲TG sequence. Fox and co-workers investigated the role of hydrogen bonding in this repair system by substituting guanine in the mismatch with inosine, 2-aminopurine, or an abasic site and by replacing the thymine in the mismatch with difluorotoluene (forming a G-F pair). The authors found that in a 50 base pair fragment containing a G-F pair in the Vsr recognition sequence, the G-F mismatch was not cleaved by the endonuclease. Given the findings in the present work, it seems plausible that the change in the asymmetry of λ angles between a G-T pair (40.5 and 66.4°) and a G-F pair (53.5 and 70.0°) is likely responsible

(62) Lukin, M.; de los Santos, C. *Chem. Rev.* **2006**, *106*, 607−686 and references therein.
(63) Kramer, B.; Kramer, W.; Fritz, H. J. *Cell* **1985**, *38*, 879−887.
(64) Fersht, A. R.; Knill-Jones, J. W.; Tsui, W. C. *J. Mol. Biol.* **1982**, *156*, 37−51.
(65) Lu, A. L.; Welsh, K.; Clark, S.; Su, S. S.; Moldrich, P. *Cold Spring Harbor Symp. Quant. Biol.* **1984**, *49*, 589−596.

for this difference in repair activity. The G-F pair does not strictly possess the phosphodiester backbone structure of an authentic mismatch (since it exhibits partial Watson−Crick nature) and is, therefore, not recognized by the repair system. This interpretation is consistent with the conclusions of Fox et al.,[26] who suggest that the Vsr endonuclease recognizes a G-T mismatch by its shape, which we more broadly define in terms of the nucleotide units rather than simply the aromatic bases and associated functional groups.

Our studies indicate that the G-F base pair formed between guanine and difluorotoluene is not identical in structure to a G-T wobble pair when placed in an identical DNA duplex. The G-F pair possesses structural features that are typical of both a mismatch and a Watson−Crick base pair. These structural differences are largely manifested in the stacking patterns and λ angles at the mispair site, while the canonical DNA structure is restored beyond one base pair on either side. These findings support the notion that subtle backbone conformations and/or dynamics likely affect recognition events that lead to DNA mismatch repair.

Acknowledgment. We thank Drs. Jeremy Kua (USD) and David Case (Scripps Research Institute) for the use of computational facilities and for helpful discussions. We thank Jeremiah Fillo for preliminary computational work on difluorotoluene nucleotide. The NMR facility at USD was established with a grant from the NSF-MRI program (0417731). Funding was provided by the Blasker-Rose-Miah Fund at the San Diego Foundation (grant to T.J.D. and D.C.T.).

Supporting Information Available: Complete chemical shift assignments for GT and GF and complete refs 54 and 55. This material is available free of charge via the Internet at http://pubs.acs.org.

JA7103608