J|A|C|S COMMUNICATIONS

Published on Web 08/17/2009

Pairing Geometry of the Hydrophobic Thymine Analogue 2,4-Difluorotoluene in Duplex DNA as Analyzed by X-ray Crystallography

Pradeep S. Pallan and Martin Egli*

Department of Biochemistry, School of Medicine, Vanderbilt University, Nashville, Tennessee 37232

Received July 11, 2009; E-mail: martin.egli@vanderbilt.edu

The 2′-deoxyribo-2,4-difluorotoluene nucleoside analogue (dF, Figure 1) was created as an isostere of 2′-deoxythymidine (dT) to investigate the role of Watson−Crick hydrogen bonds (W−C H-bonds) in DNA duplex stability and the fidelity of replication by DNA polymerases (pols).[1] Despite strengthening stacking significantly relative to dT,[2] incorporation of dF leads to a net destabilization of the duplex ($\Delta\Delta T_m = -14$ °C and $\Delta\Delta G = -3.5$ kcal mol^{-1} between the F:A- and T:A-containing DNAs).[3] Molecular dynamics (MD) simulations indicated an increased local flexibility at sites of dF incorporation,[4] although an initial NMR solution structure of a DNA duplex containing a single F:A pair provided support for similar shapes of the F:A and T:A pairs and limited conformational perturbations of the helical geometry.[3]

DNA pol I Klenow fragment (Kf exo-) inserts dATP opposite template dF with surprisingly high efficiency (V_{max}/K_m reduced 40-fold) and fidelity compared with template dT.[5] However, incorporation of dFTP opposite template dA by the same pol is inhibited more significantly relative to dTTP by comparison (>500-fold reduction in efficiency),[6] likely due to some extent to different *syn*/*anti* and sugar conformational equilibria between dF and dT. Extensive studies of the kinetics of both replicative[7] and lesion bypass (Y-class) DNA pols,[8] involving among other hydrophobic analogues dF, have led to the steric hypothesis of DNA replication; i.e., particularly the former class of pols appears to rely on shape rather than W−C H-bonding for accurate replication (reviewed in ref 1c). However, by determining crystal structures of the Y-class DNA pol Dpo4 from *S. solfataricus* in complex with DNA duplexes containing dF in the template strand, we recently found that the shapes of F:A and F:G pairs at the pol active site differ significantly from those of the canonical T:A and wobble T:G pairs.[9] The steric hypothesis of replication hinges on the assumption that dF lacks the ability to form H-bonds. This is indeed supported by semiempirical calculations that indicated distances (d_1 and d_2, Figure 1) between F and A that were increased by between 0.5 and 0.7 Å relative to the T:A pair.[10] However, to date no accurate experimental model of the F:A pair in a DNA duplex environment has been presented.

We have determined the crystal structure of a Dickerson−Drew Dodecamer (DDD) DNA duplex with a single dF nucleotide [d(CGCGAATFCGCG)]$_2$ bound to *Bacillus halodurans* ribonuclease H (*Bh*RNase H) at 1.6 Å resolution. For experimental procedures, selected crystal and refinement parameters (Table S1), and the quality of the final electron density (Figure S1), please see the Supporting Information. Although crystals of the modified DDD alone could not be grown, protein−DNA contacts in the complex are limited to the CG portion and the conformation of the central tetramer including F:A pairs is unlikely to be distorted as a result of protein binding (Figure S2). Like the structure of the complex with the native DDD,[11] the asymmetric unit of the complex with the dF-modified DDD contains two independent 12-mer strands (both duplexes are located on a dyad) and RNase H molecules.

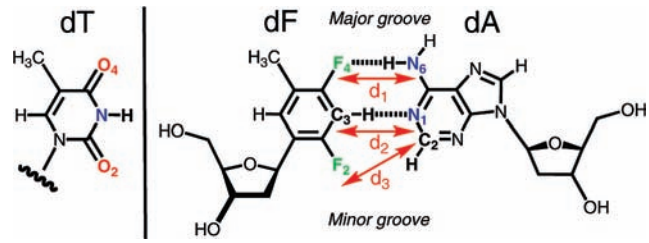

Figure 1. Structures of dT, dF, and dA and a hypothetical dF:dA pair. Putative H-bonds are dashed lines; arrows designate distances d_1 [F4(dF)···N6(dA)], d_2 [C3(dF)···N1(dA)], and d_3 [F2(dF)···C2(dA)].

Table 1. Watson−Crick H-Bond Distances d_1 and d_2 and Distance d_3 (O2/F2···C2) for T:A and F:A Pairs in Crystal Structures of B-Form DNA

B-form DNA duplex	base pair	d_1 [Å]	d_2 [Å]	d_3 [Å]	resol. [Å]	PDB code	ref.
native DDD[a]	A5:T20	2.98	2.77	3.54	1.1	436D	12a
	T8:A17	2.96	2.82	3.62			
native DDD[a]	A5:T20	3.01	2.83	3.55	1.4	355D	12b
	T8:A17	3.11	2.77	3.46			
native DDD:	F8:A17	2.95	2.81	3.57	1.8	3D0P	11
*Bh*RNase H[b]	F8:A17	2.99	2.76	3.47			
F8-DDD:	F8:A17	3.09	3.41	4.40	1.6	3I8D	this
*Bh*RNase H[b]	F8:A17	3.12	3.34	4.34			work

[a] The duplex [d(CGCG**A**AT**T**CGCG)]$_2$ (nucleotides in one strand are numbered 1−12 and 13−24 in the other) lies in a general position, and the A5:T20 and T8:A17 base pairs (bold font) exhibit different geometries. [b] The first base pair is from duplex 1 and the second from duplex 2. Both duplexes sit on a dyad, and F8:A17 and A5:F20 are symmetry equivalent.

This allowed us to analyze the geometry of two independent F:A pairs and compare them with the geometries of the corresponding T:A pairs in the native complex as well as in crystal structures at high resolution of the DDD alone (Table 1).[12]

As expected, due to replacement of N3 in dT by C3 in dF (Figure 1), d_2 is increased in the F:A relative to the T:A pair (by ∼0.5 to 0.6 Å). However, at 3.34 Å as seen in the F:A pair of duplex 2, the separation of C3(F) and N1(A) is still below the sum of the van der Waals (vdW) radii (ca. 3.7 Å). Surprisingly, d_1 is not significantly longer in F:A than in T:A (Figure 2A, B, Table 1) and well below the distance consistent with a vdW contact of ∼3.55 Å (assuming a radius of 1.35 Å for F). Even if we consider a vdW radius of F that equals that of H (1.2 Å),[13] the observed separations between F4 and NH_2 are still suggestive of an attractive interaction. At the other edge of the base pair, in the minor groove, d_3 is increased by ∼1 Å in F:A relative to T:A due to opening (Table 1).

Comparison between the conformations of the native and dF-modified duplexes in the structures of the complexes reveals only deviations near the sites of modifications (Figure 2CD). As a result of the longer d_2 distance in F:A relative to T:A, the former base

10.1021/ja905739j CCC: $40.75

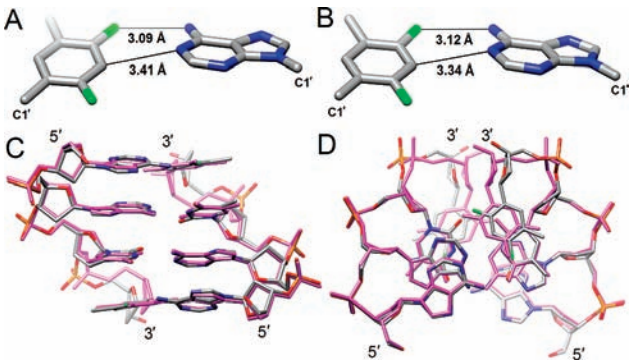

Figure 2. Geometries of the F:A pairs in (A) duplex 1 and (B) in duplex 2 with d_1 and d_2 indicated. Conformational consequences at the duplex level due to replacement of dT with dF. Superimposition of the central tetramer duplexes (AATF):(AATF) and (AATT):(AATT) from the crystal structures of the modified and native DDD in complex with *Bh*RNase H, respectively, viewed (C) along the dyad and into the major groove and (D) rotated around the horizontal by 90° and viewed along the helical axis. Duplex DNA atoms are colored gray (C; pink, native DDD), red (O), blue (N), orange (P), and green (F).

pair is stretched. This pushes the backbones outward and goes along with changes in the 10° to 20° range in backbone torsion angles ϵ and α of dF and the preceding dT. In addition to stretching, F:A pairs exhibit increased stagger (a shift of dF and dA relative to one other along the helical axis) and the aforementioned opening. However, propeller twisting that is quite pronounced in T:A pairs does not appear to be increased in F:A. Similarly, local helical twist and rise are virtually unaffected by the replacement of dT with dF (Figure 2).

In crystal structures of B-form DNA, the H-bond distance between O4 of thymine and N6 of adenine (d_1) is typically longer than that between the N3(T) and N1(A) atoms[12] (d_2, 3.04 ± 0.17 Å vs 2.83 ± 0.13 Å, respectively,[14] Table 1), and computational simulations paint a similar picture.[14] This means that the loss of the (T)N3−H···N1(A) H-bond in the F:A pair results in a considerable loss of stability relative to T:A, even if we attribute a minor stabilizing effect to the C3−H···N1 contact in F:A. On the other hand, there is a surprisingly small difference between the lengths of d_1 in the T:A and F:A pairs, and it is reasonable to postulate formation of a H-bond between F4 and N6 based on our structural data. Although we do not observe the positions of hydrogen atoms at 1.6 Å resolution, the distance between the calculated position of the N6(A) hydrogen atom and F4 amounts to ca. 2.1 Å and is thus comparable to the shortest distances found between fluorine and N−H donors in the crystal structures of small molecules.[15] The pairing mode seen here between dF and dA in B-DNA is much tighter than that between rF and rA previously analyzed in the crystal structure of an RNA duplex[16] but comparable to the pairing of rF and rG that was consistent with H-bond formation between F2(rF) and N1(rG) (min dist. = 3.03 Å).[17]

Our structural data at high resolution contradict the earlier assumption that the pairing of dF and dA does not involve H-bonding (i.e., refs 1, 10). But the structural data accumulated to date also indicate a considerable plasticity of the F:A pair, with different geometries observed in DNA here, RNA,[16,17] and at the postreplicative site of a Y-class DNA pol.[9] As far as the steric hypothesis and the reliance on shape rather than H-bonding by certain DNA pols for accurate replication are concerned, a more complicated picture is emerging. Shape and H-bonding cannot be separated readily, and steric constraints such as backbone geometry, stacking, or enzyme active sites should be considered enablers of H-bonding, in line with earlier theoretical work by Guerra and Bickelhaupt.[18] The finding here that dF engages in a H-bond to dA also raises the possibility that some DNA pols may still probe the minor groove of dF:dATP or dA:dFTP pairs at the active site with H-bonds. Data deposition: Final coordinates and structure factors have been deposited in the Protein Data Bank (http://www.rcsb.org): PDB ID code 3ID8.

Acknowledgment. This work was supported by NIH Grant R01 GM55237. Vanderbilt University is a member institution of LS-CAT at the Advanced Photon Source (Argonne, IL). Use of the APS was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38.

Supporting Information Available: Experimental procedures, Figures S1 and S2, and Table S1. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Schweitzer, B. A.; Kool, E. T. *J. Org. Chem.* **1994**, *59*, 7238–7242. (b) Kool, E. T. *Annu. Rev. Biochem.* **2002**, *71*, 191–219. (c) Kool, E. T.; Sintim, H. O. *Chem. Commun.* **2006**, 3665–3675.
(2) Guckian, K. M.; Schweitzer, B. A.; Ren, R. X.; Sheils, C.; Paris, P. L.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc.* **1996**, *118*, 8182–8183.
(3) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 954–959.
(4) Cubero, E.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2000**, *122*, 6891–6899.
(5) Moran, S.; Ren, R. X.-F.; Rumney IV, S.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057.
(6) (a) Moran, S.; Ren, R. X.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506–10511. (b) Morales, J. C.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 950–954.
(7) (a) Kim, T. W.; Delaney, J. C.; Essigmann, J. M.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15803–15808. (b) Potapova, O.; Chan, C.; DeLucia, A. M.; Helquist, S. A.; Kool, E. T.; Grindley, N. D. F.; Joyce, C. M. *Biochemistry* **2006**, *45*, 890–898. (c) Kim, T. W.; Brieba, L. G.; Ellenberger, T.; Kool, E. T. *J. Biol. Chem.* **2006**, *281*, 2289–2295.
(8) (a) Washington, M. T.; Helquist, S. A.; Kool, E. T.; Prakash, L.; Prakash, S. *Mol. Cell. Biol.* **2003**, *23*, 5107–5112. (b) Wolfle, W. T.; Washington, M. T.; Kool, E. T.; Spratt, T. E.; Helquist, S. A.; Prakash, L.; Prakash, S. *Mol. Cell. Biol.* **2005**, *25*, 7137–7143. (c) Mizukami, S.; Kim, T. W.; Helquist, S. A.; Kool, E. T. *Biochemistry* **2006**, *45*, 2772–2778.
(9) Irimia, A.; Eoff, R. L.; Pallan, P. S.; Guengerich, F. P.; Egli, M. *J. Biol. Chem.* **2007**, *282*, 36421–36433.
(10) (a) Santhosh, C.; Mishra, P. C. *Int. J. Quantum Chem.* **1998**, *68*, 351–355. (b) Wang, X.; Houk, K. N. *Chem. Commun.* **1998**, 2631–2632.
(11) Pallan, P. S.; Egli, M. *Cell Cycle* **2008**, *7*, 2562–2569.
(12) (a) Tereshko, V.; Minasov, G.; Egli, M. *J. Am. Chem. Soc.* **1999**, *121*, 470–471. (b) Shui, X.; McFail-Isom, L.; Hu, G. G.; Williams, L. D. *Biochemistry* **1998**, *37*, 8341–8355.
(13) Dunitz, J. D.; Schweizer, W. B. *Chem.—Eur. J.* **2006**, *12*, 6804–6815.
(14) Srinivasan, A. R.; Sauers, R. R.; Fenley, M. O.; Boschitsch, A. H.; Matsumoto, A.; Colasanti, A. V.; Olson, W. K. *Biophys. Rev.* **2009**, *1*, 13–20.
(15) (a) Howard, J. A. K.; Hoy, V. J.; O'Hagan, D.; Smith, G. T. *Tetrahedron* **1996**, *52*, 12613–12622. (b) Dunitz, J. D. *ChemBioChem* **2004**, *5*, 614–621.
(16) Xia, J.; Noronha, A.; Li, F.; Rajeev, K. G.; Akinc, A.; Braich, R.; Egli, M.; Manoharan, M. *Am. Chem. Soc. Chem. Biol.* **2006**, *1*, 176–183.
(17) Li, F.; Pallan, P. S.; Maier, M. A.; Rajeev, K. G.; Mathieu, S. L.; Kreutz, C.; Fan, Y.; Sanghvi, J.; Micura, R.; Rozners, E.; Manoharan, M.; Egli, M. *Nucleic Acids Res.* **2007**, *35*, 6424–6438.
(18) Guerra, C. F.; Bickelhaupt, F. M. *Angew. Chem., Int. Ed.* **2002**, *41*, 2092–2095.

JA905739J