The Steric Hypothesis for DNA Replication and Fluorine Hydrogen Bonding Revisited in Light of Structural Data

MARTIN EGLI*

Department of Biochemistry, School of Medicine, Vanderbilt University, Nashville, Tennessee 37232, United States
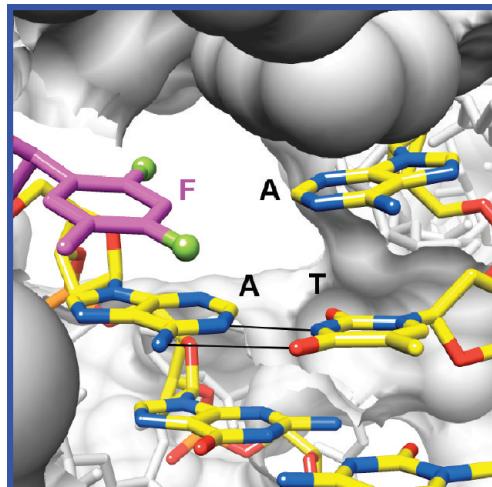
RECEIVED ON NOVEMBER 23, 2011

CONSPECTUS

In DNA, bases pair in a molecular interaction that is both highly predictable and exquisitely specific. Therefore researchers have generally believed that the insertion of the matching nucleotide opposite a template base by DNA polymerases (pols) required Watson−Crick (W−C) hydrogen bond formation. However pioneering work by Kool and co-workers using hydrophobic base analogs such as the thymine (T) isostere 2,4-difluorotoluene (F) showed that shape rather than H-bonding served as the primary source of specificity in DNA replication by certain pols. This steric hypothesis for DNA replication has gained popularity, perhaps discouraging further experimental studies to address potential limitations of this new idea.

The idea that shape trumps H-bonding in terms of pol selectivity largely hinges on the belief that fluorine is a poor H-bond acceptor. However, the shape complementarity model was embraced in the absence of any detailed structural data for match (F:A) and mismatch pairs (F:G, F:C, F:T) in DNA duplexes or at active sites of pols. Although the F and T nucleosides are roughly isosteric, it is unclear whether F:A and T:A pairs exhibit similar geometries. If the F:A pair is devoid of H-bonding, it will be notably wider than a T:A pair. Because shape and size and H-bonding are intimately related, it may not be possible to separate these two properties. Thus the geometries of an isolated F:A pair in water may differ considerably from an F:A pair embedded in a stretch of duplex DNA, at the tight active site of an A-family replicative pol, or within the spacious active site of a Y-family translesion pol. The shape complementarity model may have more significance for pol accuracy than efficiency: this model appears to be most relevant for replicative pols that use specific residues to probe the identity of the nascent base pair from the minor groove side. However, researchers have not fully considered the importance of such interactions that include H-bonds compared with W−C H-bonds in terms of pol fidelity and the shape complementarity model.
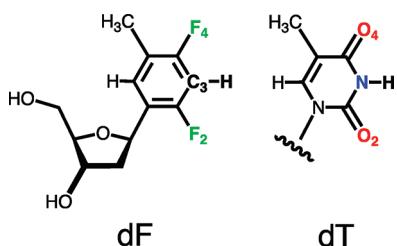
This Account revisits the steric hypothesis for DNA replication in light of recent structural data and discusses the role of fluorine as an H-bond acceptor. Over the last 5 years, crystal structures have emerged for nucleic acid duplexes with F paired opposite to natural bases or located at the active sites of DNA pols. These data permit a more nuanced understanding of the role of shape in DNA replication and the capacity of fluorine to form H-bonds. These studies and additional research involving RNA or other fluorine-containing nucleoside analogs within duplexes indicate that fluorine engages in H-bonding in many cases. Although T and F are isosteric at the nucleoside level, replacement of a natural base by F in pairs often changes their shapes and sizes, and dF in DNA behaves differently from rF in RNA. Similarly, the pairing geometries observed for F and T opposite dATP, dGTP, dTTP, or dCTP and their H-bonding patterns at the active site of a replicative pol differ considerably.

Introduction

The finding that certain DNA polymerases (pols) insert dATP opposite a 2,4-difluorotoluene nucleoside analog (F, Figure 1) quite efficiently and accurately relative to template T is counterintuitive.[1−3] Although we have known for a long time that enzymes rely to various degrees on shape and size differences to select their substrates,[4,5] Watson−Crick hydrogen bonding (W−C H-bonding) is

FIGURE 1. Structures of dT and dF (or F).

crucial for the pairing specificity and stability of DNA. That shape should ultimately be the major or sole determinant of replication fidelity as put forth in the so-called *Steric Hypothesis for DNA Replication*[6−9] is therefore startling. Moreover, the shape complementarity model hinges largely on the assumption that despite its high electronegativity, organic fluorine is a poor H-bond acceptor.[10,11] While this verdict is consistent with the environments of fluorine in small-molecule crystal structures, a host of physical−chemical data, and the results of computational simulations, there are numerous reports of fluorine affecting crystal packing and protein−ligand and DNA−ligand interactions directly or indirectly despite the weakness of individual C−H···F, C−F···F, C−F···N, or C−F···π contributions (refs 12−15 and references therein). Although there is a priori no reason to believe that difluorotoluene violates the rule that fluorine hardly ever accepts H-bonds,[11] it is obvious that without actual structural data for the analog in the relevant contexts, that is, in a DNA duplex or at a pol active site, we cannot rule out that the supposedly hydrophobic T isostere might occasionally engage in H-bonding interactions. Shape, despite being a key player in enzyme function and activity, is by no means the only determinant of substrate specificity.

Over the last 5 years, my laboratory has determined crystal structures of DNA and RNA oligonucleotides that contain F or other fluorine-modified nucleoside analogs. We as well as others have also analyzed crystal structures of pol−DNA complexes with F in the template strand and at the active site. These data permit a more differentiated interpretation of the role of shape in DNA replication. In combination with experimental stability data, the structures of duplexes containing fluorine-bearing residues support the view that fluorine can form energetically favorable H-bonds. Despite the growing database of structures containing F, predictions regarding the geometry of and contributions of fluorine to H-bonds in dF/rF match and mismatch pairs remain difficult. What is clear is that both geometry and fluorine H-bonding are strongly influenced by the particular environment.

The Hypothesis

Looking back at a decade of debate regarding the hydrophobicity of the difluorotoluene analog, Sintim and Kool remained unconvinced of the existence of N−H···F or C−H···F H-bonds.[8] A host of observations were offered in support of the strictly hydrophobic nature of F or dF. For example, dF partitions strongly to octanol in mixtures with water, unlike dT [log P(dF) = 0.78; log P(dT) = −1.1]. In the crystal structure of dF, F_2, C_3H, and F_4 appear not to form H-bonds, whereas in the structure of dT, the corresponding O_2, N_3H, and O_4 do. The dipole moments of F and T differ strongly (1.84 vs 4.19 D, respectively), as do the atomic charges of C_3(F) and N_3(T) (0.074 vs 0.203, respectively) and F_4(F) and O_4(T) (−0.079 vs −0.307, respectively). Incorporation of a single F in place of T into DNA opposite A destabilizes the duplex significantly; the T_m as assessed by UV melting drops by some 15 °C.[16] However, a 12mer DNA duplex with an F:A pair is more stable than duplexes with F:G, F:C, or F:T mismatch pairs (by between 2 and 3 °C). Perhaps this reflects a better shape correspondence and more optimal stacking with neighboring pairs as a result or a residual electrostatic contribution that can be expected to be more favorable for a matching pair. Interestingly, when the relative stabilities of U, F, and the toluyl (*To*) C-nucleoside analog paired with either A, G, U, or C inside duplex RNA were compared, *To* pairs are clearly more destabilizing than F pairs (i.e., U:A, 37.8 °C; F:A, 27.4 °C; *To*:A, 23.0 °C).[17] This observation too leaves open the possibility that F pairing entails a residual H-bonding contribution, although it is also known that F furnishes improved stacking in DNA compared with both T and a benzyl C-nucleoside analog.[18]

At the level of the nucleoside, there is good evidence that dF is an isostere of dT. The C=O and C−F bond lengths are not too different. Both dF and dT favor the *anti* glycosidic conformation and the preferred sugar conformation with both lies in the Southern range (90% and 70% *South* for dF and dT, respectively). The NMR structure of a DNA duplex with an incorporated F:A pair appeared to be largely unchanged relative to B-form DNA,[19] further bolstering the impression that dF and dT and perhaps the F:A and T:A pairs are of the same shape. However, the structure itself did not provide a detailed look at the F:A pair. Rather, the structural data simply support the notion that F:A does not significantly perturb the DNA duplex, that is, by F or A or both adopting an extra-helical orientation. So the shapes may be similar, but the sizes (width of the base pair, that is, glycosidic C1′···C1′ distance) could nevertheless differ substantially between the
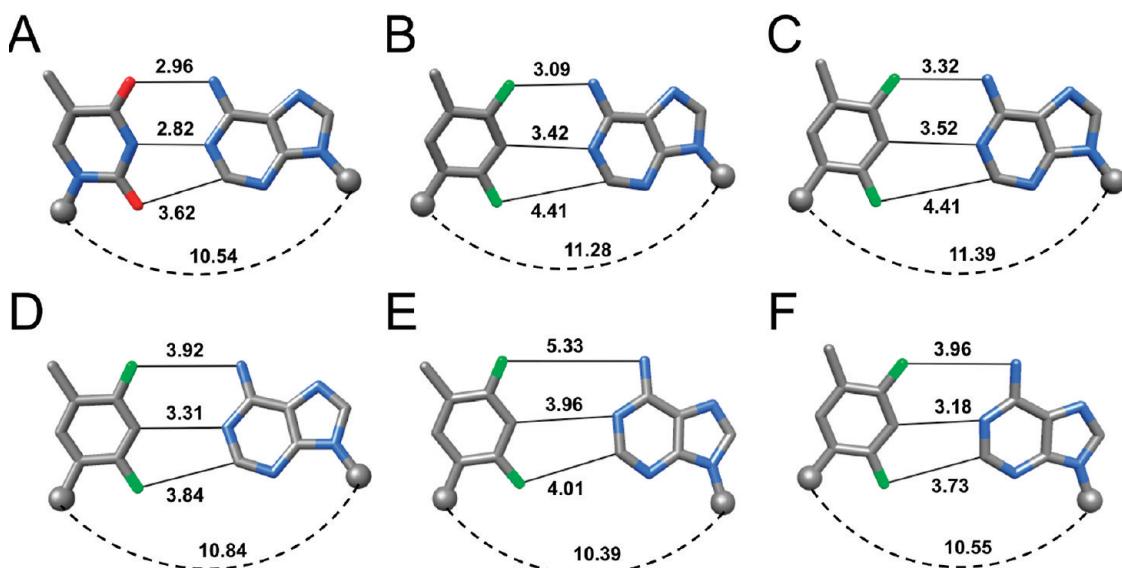
FIGURE 2. Geometries of F:A pairs: (A) reference T:A pair (DDD, CGCGAATTCGCG; PDB ID 436D);[32] (B) F:A in a modified DDD (CGCGAATFCGCG; PDB ID 3I8D);[30] (C) computational model of F:A;[20] (D) rF:rA in duplex RNA (rCGCFAAUUAGCG; PDB ID 2G92);[33] (E) F:A at the active site of the Y-family pol Dpo4 (PDB ID 2VA2);[25] (F) F:A at the active site of the B-family pol RB69 (PDB ID 3QER).[43] Distances are in Å; C1′ atoms are shown as spheres, and the spacing between them is indicated by a dashed arc.

two. In fact, ab initio calculations and force field modeling of the F:A pair in duplex DNA were not suggestive of H-bonding and the resulting $F_4 \cdots H-N_6$ and $C_3-H \cdots N_1$ distances of 3.32 and 3.52 Å, respectively,[20] were around 0.5 Å longer than the corresponding distances in a T:A pair (Figure 2). Thus, the inability of fluorine in dF to accept H-bonds and the isosteric nature of the analog relative to T are two cornerstones of the steric hypothesis for DNA replication.

The central pillar of the hypothesis that electrostatic effects seem to play a minor role in accurate replication, ultimately triggering more extensive investigations into the importance of shape, is of course provided by the efficient and preferred (as judged by steady-state kinetics, k_{cat}/K_m) incorporation by *Escherichia coli* DNA polymerase I Klenow fragment (Kf) of dATP opposite template F relative to dATP insertion opposite T,[1] as well as that of insertion of dFTP opposite A relative to dTTP opposite A.[2] There is an asymmetry in the decrease of the efficiency in that F in the template is better tolerated than F as the incoming nucleotide triphosphate. And with Kf and other replicative pols, efficiency is more affected than fidelity.[2,9,21] The asymmetry between dATP insertion opposite template F and dFTP insertion opposite template A becomes much more pronounced when pre-steady-state kinetics are considered. Thus, the reduction in efficiency (k_{cat}/K_d) is around 30-fold in the former case but much more pronounced (ca. 1000-fold) for dFTP vs dTTP as the incoming nucleotide.[22] Similar observations with respect to efficiency and accuracy

were made with the human mitochondrial pol.[23] Again, these data indicate the importance of H-bonding for efficient primer extension even with pols featuring tight active sites.

Compared with A- and B-family pols, the tolerance of Y-family translesion pols toward F in the template or dFTP as the incoming nucleotide is much more limited. The archaeal Dbh[22] and Dpo4[24,25] pols, as well as yeast pol η[26] and human pol κ,[27] were found to be dependent on H-bonding for efficient and accurate replication. Thus, it appears that only certain pols (i.e., those belonging to the A- and B-families) with tight active sites and amino acids probing the minor groove of the replicative and postreplicative base pairs[22] are capable of accurate and relatively efficient replication, by selecting incoming and templating nucleotide pairs of optimal shape and size. Interestingly, Dpo4 (Y-family) and Kf (A-family) exhibit similar preferences vis-à-vis the gradually increasing sizes of base pairs between A and various thymine analogs (toluene, 2,4-difluorotoluene, 2,4-dichlorotoluene, 2,4-dibromotoluene, 2,4-diiodotoluene).[24] For both pols, the maximum efficiency is observed with the dichloro analog (either inside the template or as the incoming nucleotide triphosphate). However, given the important role attributed to shape and size, it is puzzling that Kf's efficiencies of incorporating either dFTP or the 2′-deoxy diiodotoluene triphosphate analog (d[DIT]TP) opposite A are virtually identical. This indicates that steric effects alone cannot account for the overall selectivity of Kf.

F Pairing in Duplex DNA

The B-form DNA duplex [d(CGCGAATTCGCG)]$_2$, the so-called Dickerson−Drew dodecamer (DDD), has long served as a crystallographic model system to study the sequence dependence of conformation, bending, hydration, and metal ion binding (ref 29 and references therein). We synthesized a modified DDD with F inserted in place of T8, d(CGCGAATFCGCG) and determined its structure bound to RNase H from *Bacillus halodurans* (*Bh*RNase H) since crystals could not be grown for the duplex alone.[30] The protein in the complex crystals merely serves as a scaffold and interacts with the G-tracts at both ends of the DDD.[31] Thus the geometry of the central AATT (AATF) tract is not perturbed as a result of the presence of RNase H. At 1.6 Å, the structure is of sufficiently high resolution to allow accurate determination of interatomic distances (see Supporting Information). In the two independent F:A pairs observed in the structure, the F$_4$(F)$\cdots$N$_6$(A) distances are 3.09 and 3.12 Å (Figure 2B). These are not too different from the corresponding O$_4$(T)$\cdots$N$_6$(A) distances in structures of the native DDD or its complex with *Bh*RNase H, which vary between 2.96 and 3.11 Å (Figure 2A).[30,32] At just slightly above 3 Å, the experimentally determined distances involving fluorine are significantly shorter than those seen in computational models of F:A pairs in which the minimal separation reported amounts to 3.32 Å (Figure 2C).[20] However, most of the modeled F:A pairs in DNA or RNA feature F$_4$(F)$\cdots$N$_6$(A) separations that are more consistent with the sum of the van der Waals (vdW) radii of the interacting atoms (ca. 3.6 Å).[17,30] Because the distance is around 3.1 Å in the crystal structure and thus about 15% shorter than the above sum of vdW radii, we can basically rule out a mere van der Waals contact between fluorine and the amine in the DDD F:A pairs. If we assume a bond length of around 1 Å for N−H, then the 3.1 Å F$\cdots$N separation in the DDD F:A pairs matches the shortest F$\cdots$H−N interactions in crystal structures of small molecules, which were found to be around 2.1 Å.[10] H-bond participation by organic fluorine may indeed be rare, but it is hard to avoid the impression that electrostatics have something to do with the observed configuration of F:A pairs in the DDD. Although incorporation of F in place of T into DNA destabilizes the duplex significantly, DNAs with F:A pairs melt at higher temperatures than those with F:G, F:C, or F:T mismatch pairs,[16] as pointed out in the Introduction. Clearly, the stabilization afforded by a putative C−F$\cdots$H−N H-bond is limited compared with the standard C=O$\cdots$H−N H-bond. Either favorable stacking or residual H-bonding or both are likely at the origin of the higher stability of F:A relative to the mismatch pairs.

F Pairing in Duplex RNA

The effects of incorporation of the ribonucleotide analog of F, rF, into small interfering RNA molecules (siRNAs) on their activity have been studied in some detail.[33−35] The modulation of the siRNA activity upon replacement of rU (U) by rF in the antisense (guide) strand was assayed in HeLa cells *in vitro*.[33,34] When placed at the 5′-terminal end or opposite A in the interior of the guide strand, rF was quite well tolerated. Even when rF was positioned adjacent to the Ago2 cleavage site, measurements of the luciferase luminescence revealed that the modified siRNA duplex maintains considerable activity as long as the pairing partner of rF is A. Thus, rF:A and C:A at the cleavage site affected the activity to a similar degree, and these combinations were better tolerated than A:A or G:A mismatches, which strongly impaired activity.[33] A related study that assessed the interference of siRNAs with rF in place of U at 11 different positions of the guide strand demonstrated that activity and thermodynamic stability of the individual duplexes correlated quite well.[34] However, an siRNA with rF at a central location of the guide strand retained high activity despite being strongly destabilizing. Interestingly, the crystal structure of the RNA duplex [r(CGCFAAUUAGCG)]$_2$ at 1.6 Å resolution revealed a configuration of the rF:A pair with a separation between F$_4$ and N$_6$ that was considerably larger than that in the corresponding pair in the DDD (Figure 2D).[33] At almost 4 Å, the distance exceeds the sum of van der Waals radii by some 10%, indicating a rather loose association of F and A in RNA and hinting at a potentially considerable range of the F:A geometry inside canonical duplexes. This finding is consistent with the results of molecular dynamics (MD) simulations for an RNA containing a central rF:A pair.[36] Compared with a reference duplex with U:A, the modified duplex exhibited significantly larger conformational fluctuations in the central portion, with frequent base pair openings and shearing.

Whereas single rF:A pairs in the central portion of the guide strand resulted in an only marginally reduced activity compared with native siRNA, three consecutive rF:A pairs covering the same region had a detrimental effect on the activity.[35] Evidently the softer core of the latter construct is not tolerated by the RNA-induced silencing complex (RISC) and perhaps hampers cleavage catalyzed by Ago2. The activity of an siRNA duplex with a single U:G mismatch in the guide strand adjacent to the cleavage site was as low as that of the above duplex with three rF:A mismatches.
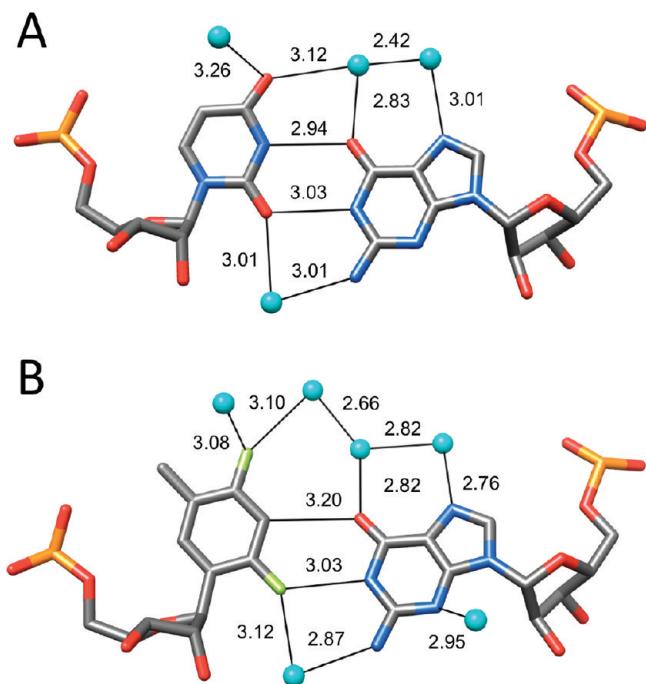
FIGURE 3. Geometry and hydration of RNA (A) U:G (rGGGGCUA:rUAG-CUCC; PDB ID 434D)[38] and (B) rF:G (rCGCFAAUUGGCG; PDB ID 2Q1O)[35] pairs. Water molecules are cyan spheres, and distances are in Å.

And replacing the same U:G by rF:G or placing the rF:G mismatch two nucleotides upstream in the guide strand abolished the activity altogether.[35] When we analyzed the rF:G mismatch in the crystal structure of the duplex [r(CGCFAAUUGGCG)]$_2$ at 1.1 Å resolution, we found to our surprise that its geometry mimicked that of the native U:G wobble pair (Figure 3). Given the weak association between rF and A in the same sequence context, this is rather puzzling. Not only does the F$_2$ atom in rF form a close contact with the imino N$_1$—H from G (3.03 Å) in the minor groove, but the 3.2 Å distance between C$_3$—H (rF) and O$_6$(G) is also suggestive of a C—H···O H-bond. By comparison, the C$_3$—H···N1 distances in the dF:A (DDD; Figure 2B) and rF:A pairs (RNA; Figure 2D) are longer and not compatible with a significant role in pairing stability.

The mimicry of U:G by rF:G is not limited to the nucleotides but extends to the hydration of the pairs, with water molecules engaged in close contacts to F$_2$ and F$_4$ atoms (Figure 3B). Rather than focusing solely on the tight spacing between donor atoms and fluorine, which is consistent with H-bond formation, the most convincing argument for electrostatics (not stacking) underlying the geometry of the rF:G pairs in our crystal structure is the displacement of G into the minor groove. Why, other than for favorable electrostatics, that is, H-bonding, would rF and G slide along each other to
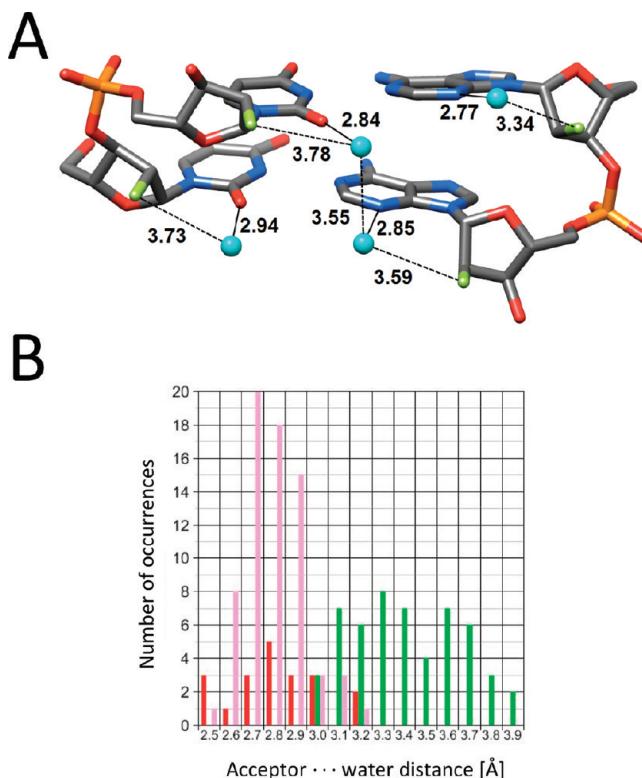

FIGURE 4. Fluorine is a poor acceptor in the minor groove of FRNA: (A) example of the water structure in [f(CGAAUUCG)]$_2$ (PDB code 3P4A);[39] (B) comparison between distances to water molecules by O2′ (red), O$_2$ (pyrimidine)/N$_3$ (purine) (pink), and F2′ (green) in the minor grooves of FRNA or mixed FRNA/RNA duplexes.[40]

assume a relative orientation that is basically identical with that seen for U and G (Figure 3)?

An MD simulation of F and G in B-DNA resulted in a loosely associated pair in a pseudo-W—C configuration with F$_2$(F) spaced at about 3.5 Å from N$_2$(G) in the minor groove,[37] inconsistent with our crystallographic results in the context of RNA. That rF in RNA behaves quite differently from F in DNA is supported not only by the tighter pairing between mismatched rF and G (Figure 3B) compared with the rF:A matching pair (Figure 2D) but also by divergent effects of F incorporation into DNA and RNA on the thermodynamic stabilities of the respective duplexes. For example, we found that RNA with an rF:G mismatch melts at a slightly higher temperature than the same duplex with rF:A.[35] This is noteworthy because the native duplex with U:A has higher stability than the corresponding duplex with U:G. In addition, osmotic stress experiments demonstrated that significantly more water molecules are released upon melting of an RNA duplex with rF:G relative to the same duplex with rF:A, in line with our observation that the former base pair is surrounded by more water molecules in the crystal.[33,35]
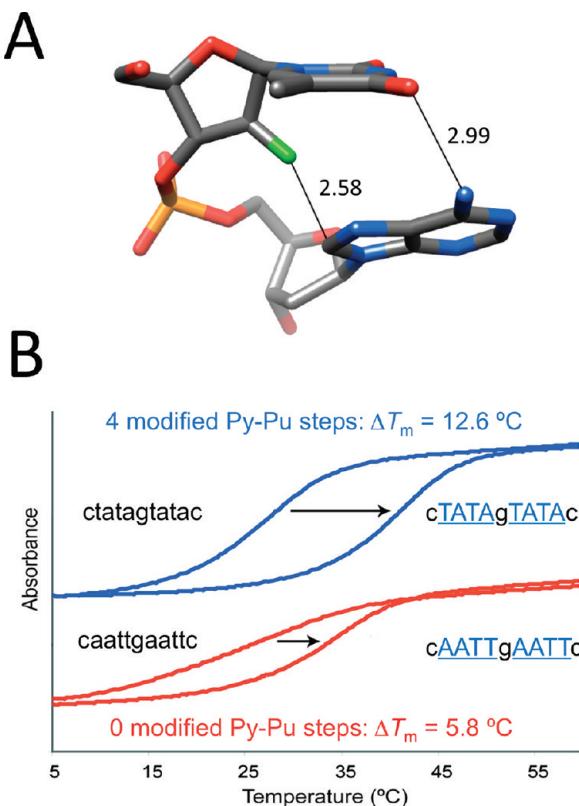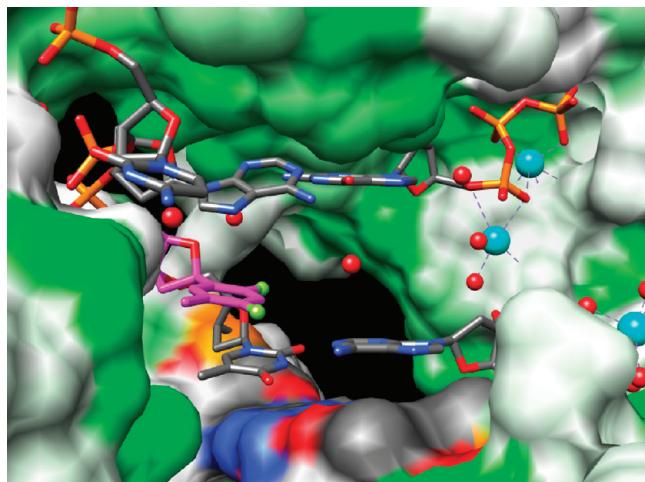
FIGURE 6. The roomy active site of the Dpo4 Y-family pol can accommodate two template bases (A and F in this so-called type II complex; PDB ID 2VA3).[25] H-bonding is required between template base and incoming dNTP for efficient bypass. The pol surface is colored according to hydrophobicity (min = white; max = green), F carbons are magenta, and Ca^{2+} ions and water molecules are cyan and red spheres, respectively.

FIGURE 5. Stabilizing effect of C−H···F interactions in FANA-modified DNAs: (A) tight contact between FANA-T 2′-fluorine and C_8−H of dA at a (faT)p(dA) Py−Pu step in an A-form duplex (PDB code 2FIL);[41] (B) incorporation of FANA residues (capital font in blue) into an oligo-2′-deoxynucleotide (small font) at Py−Pu steps leads to increased stability of the duplex with complementary RNA (UV melting assay) compared with a mixed FANA−DNA strand of identical composition but lacking a Py−Pu step.[42]

Stabilizing H-bond Interactions by Fluorine

The crystallographic data for F incorporated into DNA and RNA, although still somewhat limited, demonstrate that fluorine does not behave in a uniform manner. Rather, it is a sort of polar hydrophobe whose interactions appear to be influenced by the particular environment. Structural and thermodynamic studies relating to 2′-deoxy-2′-fluoro-ribonucleic acid (FRNA) and 2′-deoxy-2′-fluoro-arabinonucleic acid (FANA) provide further support for this dichotomy. Crystal structures of FRNA duplexes support the notion that fluorine replacing the ribose 2′-hydroxyl group is hydrophobic and does not engage in H-bond formation (Figure 4A).[39,40] Fluorine polarizes the nucleobase and may strengthen W−C H-bonds but, unlike 2′-OH, cannot act as a bridgehead for water molecules linking paired strands across the minor groove. This is evident from a comparison of the distances between water molecules and 2′-F, 2′-OH, or base functions such as pyrimidine O_2 and purine N_3 (Figure 4B). Therefore,

fluorine exposed to water is no match for H-bond acceptors such as hydroxyl, ketones, or imines. Conversely, fluorine in the arabino configuration in FANA is sequestered from solvent inside the duplex and was found to engage in a tight contact with C_8−H of the 3′-adjacent adenine in the crystal structure of an A-DNA duplex with incorporated FANA-Ts (Figure 5A).[41] A short distance between nonbonded atoms alone, even if these constitute a potential donor−acceptor pair, cannot be taken as evidence for a stabilizing interaction. However, in the case of the F2′···H−C contact at FANA intrastrand pyrimidine−purine steps, there is experimental evidence from standard UV melts that such fluorine (pseudo) H-bonds are energetically favorable (Figure 5B).[42] The FRNA and FANA examples provide a powerful demonstration of altered fluorine behavior as a consequence of an epimeric change and different environments.

F Pairing at a Y-Family Pol Active Site

Y-family translesion pols possess spacious active sites (Figure 6), consistent with their roles of inserting dNTPs opposite or extending from adducted bases.[22−27] With the Dpo4 pol from *Sulfolobus solfataricus* insertion of dATP opposite F was ca. 5000-fold less efficient than insertion opposite T, as measured by steady-state kinetics (k_{cat}/K_m).[25] However, in the steady state, Dpo4 exhibited a robust ability to discriminate between inserting either dATP or dGTP opposite F, as the efficiency in the latter case was reduced about 200-fold. Consistent with the >5000-fold reduced

efficiency in the steady state, no product (F:A) was formed in the pre-steady state, even at high dATP concentrations.[25] Also in the pre-steady state, extension following an F:A pair was moderately inhibited relative to T:A, but no extension at all was seen following F:G.
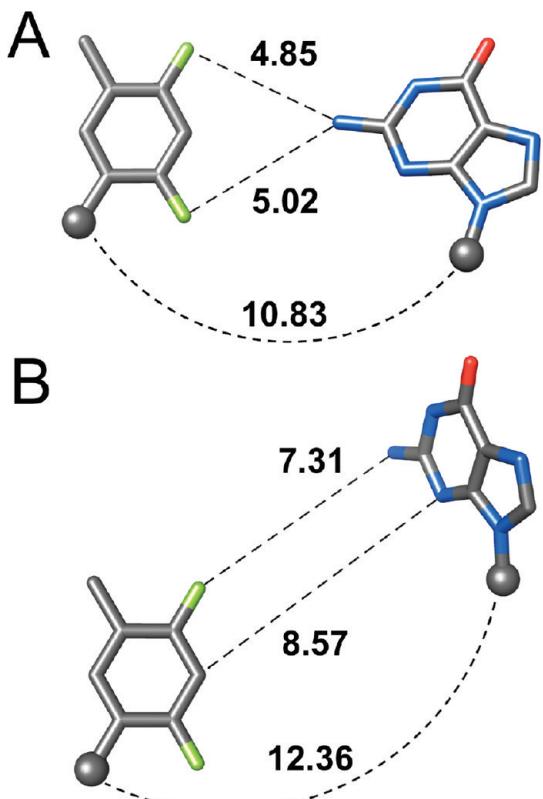

FIGURE 7. Geometry of F:G at the active site of the Y-family pol Dpo4 (PDB ID 2V9W).[25]

The crystal structure of Dpo4 with F:A at the active site revealed a loose association between the two, with F_4(F) and N_6(A) separated by >5 Å (Figure 2E).[25] However, despite this large separation, the distance between C1′ atoms was basically identical to that in a native T:A pair (Figure 2). The geometry of the F:A pair seen here differs drastically from the tightly associated F and A in the core of the DDD (Figure 2B),[30] indicating once again that the environment crucially affects shape and size of base pairs involving F. This is particularly evident in the structure of a ternary complex of Dpo4 with F lodged opposite G at the active site that allowed two independent observations of this pairing (Figure 7).[25] Relative to F:A, the exocyclic N_2 amino group of G pushes the difluorotoluene moiety further away (Figures 2E and 7A). Since neither G nor A forms H-bonds to F at the Dpo4 active site, the pol ironically may rely solely on steric factors to discriminate between the two. In the second complex, G was completely ejected out of the active site (Figure 7B). It is clear that such a misalignment will prevent a proper line up between the 3′-terminal hydroxyl group of the primer and the α-phosphate of the incoming nucleotide (i.e., dGTP), required for efficient replication.

F Pairing at a B-Family Pol Active Site

A- and B-family pols possess tight active sites, consistent with their ability to replicate DNA with high accuracy and efficiency (Figure 8). Thus, such pol active sites may select for standard size W−C pairs between natural purines and pyrimidines. Crystal structures for ternary complexes of the RB69 B-family pol (the L561A/S565G/Y567A triple mutant) with template F opposite dATP, dTTP, dGTP, or dCTP were
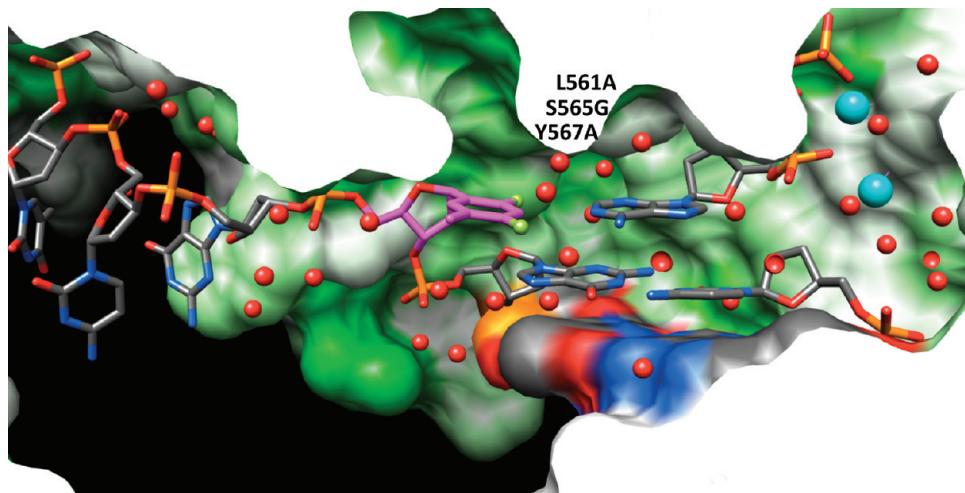

FIGURE 8. The tight active site of the RB69 B-family pol with template F opposite dATP (PDB ID 3QER).[43] Note that portions of the pol surface are sliced away to allow a view into the active site from the major groove side and that access to the minor groove is blocked compared with Dpo4 (Figure 6; please see that figure for color codes).

reported very recently.[43] In the F:dATP complex, F and A were found to be more closely spaced (Figure 2F) relative to the situation in Dpo4 (Figure 2E). However, the distance of ca. 4 Å between F_4(F) and N_6(A) is inconsistent with H-bonding. The geometry of the F:A pair in the RB69 structure actually bears close resemblance to that of F:A in duplex RNA (Figure 2D), at once invalidating the authors' suggestion that the geometries of pairs involving F at pol active sites and within isolated duplexes differ as a result of backbone constraints in the latter. As far as the steric hypothesis is concerned the F:A arrangement in RB69 is not a close mimic of T:A (i.e., Figure 2A). With regard to the ability of fluorine to participate in H-bonds, the structure of the complex with F:A at the RB69 active site stands in contrast to our observations for F:A in the DDD (Figure 2B).[30]

That the tight active site of the B-family pol can be conducive to fluorine H-bonding was demonstrated by the crystal structure of RB69 with mismatched F:dTTP (Figure 9A). Not only did F_4(F) engage in a contact <3 Å with N_3(T), but both Cα and amide N−H of glycine 568 that seals the minor groove were located quite close to the F_2 atom (similar in the other structures). However, these latter contacts involving F_2 as well as those between F_4 and water molecules wedged into the active site do not by themselves provide strong support for fluorine H-bonding (Figure 9A). Moreover, the mutations in the RB69 pol studied affect amino acids L561, S565, and Y567 that line the minor groove (see Figure 8 for orientation). Potential effects of the latter on the F pairing mode can therefore not be fully evaluated. By comparison, F and dCTP lie far apart in the active site (Figure 9B), providing a rationale for the slightly

more efficient incorporation of T ($k_{pol}/K_d = 2.8 \times 10^{-2} \mu M^{-1} s^{-1}$) over C ($k_{pol}/K_d = 1.3 \times 10^{-2} \mu M^{-1} s^{-1}$) opposite F ($k_{pol}/K_d$, F:dATP = 1.6 $\mu M^{-1} s^{-1}$). The preferred incorporation of dATP opposite template F is likely a consequence of the overall fit within the active site and the enhanced stacking of the incoming purine base relative to T and C.
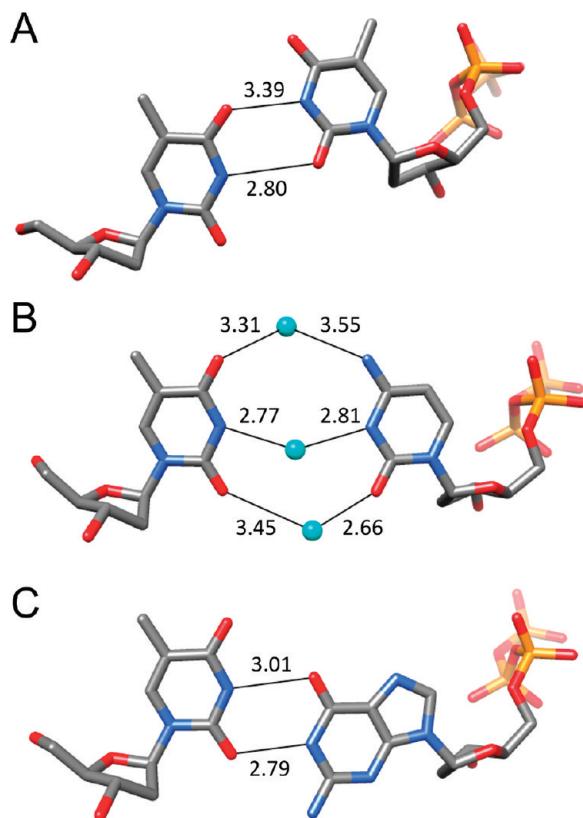

FIGURE 10. Geometry of (A) T:dTTP (PDB ID 3QET), (B) T:dCTP (PDB ID 3QEV), and (C) T:dGTP (PDB ID 3QEX) at the RB69 active site.[43]
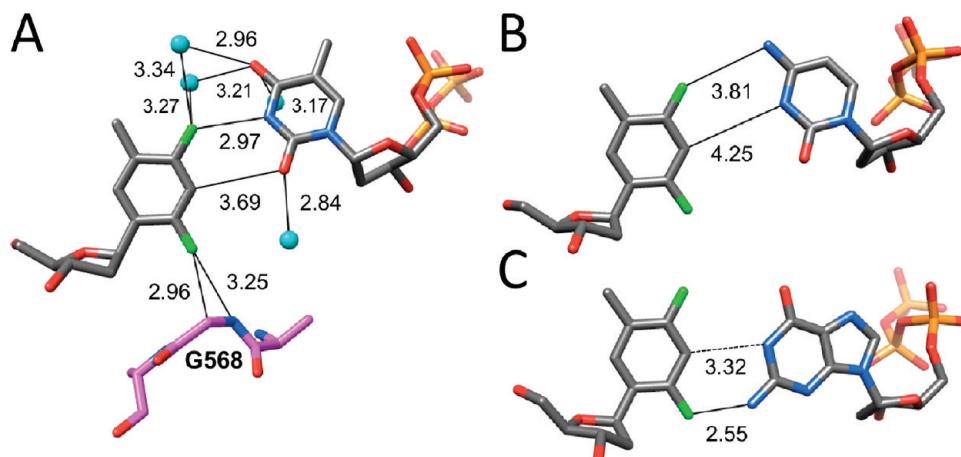

FIGURE 9. Geometry of (A) F:dTTP (PDB ID 3QEP), (B) F:dCTP (PDB ID 3QEI), and (C) F:dGTP (PDB ID 3QES) at the active site of the B-family pol RB69.[43] Pol residues A567, G568, and A569 probing the nascent base pair from the minor groove side are depicted with carbon atoms colored in magenta (similar for F:dCTP and F:dGTP) and selected water molecules are cyan (F:dTTP).
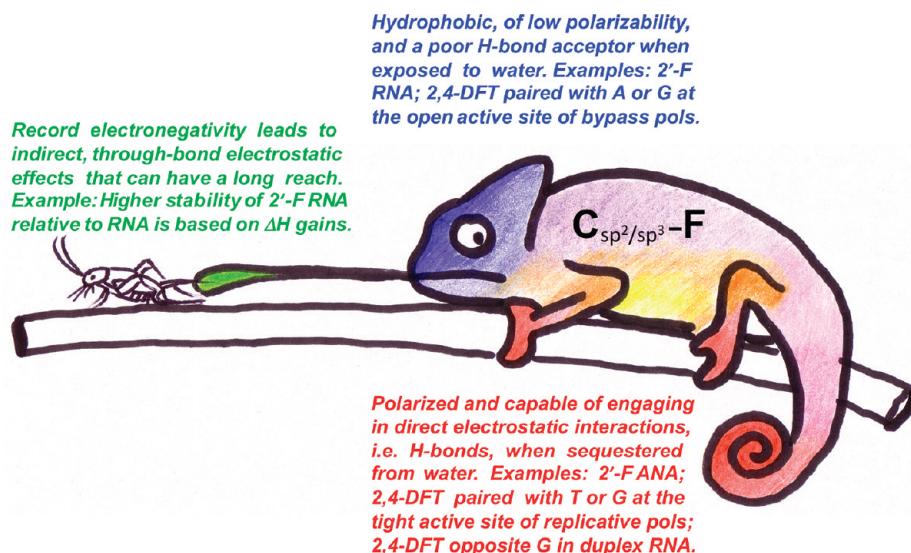
FIGURE 11. Organic fluorine is a chemical chameleon, its "color" (electronic preference) depending on the local environment.

Insertion of dGTP opposite F proceeds with the lowest efficiency ($k_{pol}/K_d = 9 \times 10^{-3}$ $\mu M^{-1}s^{-1}$), and the structure reveals a buckled orientation of G (Figure 9C) with diminished stacking on the penultimate base pair. However, the $(F)F_2\cdots H-N_2(G)$ and $(F)C_3-H\cdots H-N_1(G)$ distances are unrealistically short in the structure (Figure 9C) and are well below what can be expected for a fluorine H-bond and a $C-H\cdots H-N$ van der Waals interaction, respectively.

Looking at these four base pairs involving F and incoming dNTPs at the RB69 active site (Figures 2F, 9) and comparing them to the corresponding pairs with T (Figures 2A, 10) makes it clear that template F and T behave quite differently. Thus, it is hard to avoid the impression that one may have compared apples and oranges by simply invoking sterics as the source of the ability of replicative pols to preferentially incorporate dATP opposite F.

Conclusions

Crystal structures of DNAs, RNAs, and pol–DNA complexes containing dF (rF in RNA) or other fluorine-modified nucleosides (FRNA, FANA) determined in recent years demonstrate that the standard view of fluorine as essentially hydrophobic and unable to participate in H-bonds belies a more complex behavior that appears to be influenced by the molecular environment and involves participation in H-bonds in many cases (Figure 11). Fluorine in a polar environment, that is, exposed to water as in FRNA (Figure 4), does not engage in H-bonding. By contrast, when sequestered from solvent, that is, in the core of duplexes (Figures 2B, 3B, 5A) or at active sites of replicative pols (Figure 9A), fluorine can participate in H-bonds. However, such H-bonds provide only weak

constraints as can be seen from the wide range of geometries exhibited by F:A pairs (Figure 2) or the divergent geometries of F:G in duplex RNA (Figure 3B), at the active site of the Dpo4 pol (Figure 7), or at the active site of the RB69 pol (Figure 9C). Based on the currently available data, F behaves quite differently in DNA and RNA. Clearly, a key assumption of the steric hypothesis for DNA replication, namely, that the F analog is unable to engage in H-bonds, is no longer tenable. Further, the pairing geometries of F (Figure 9) and T (Figure 10) at a pol active site are quite different, indicating that F is a rather poor shape mimic of T even when it does engage in H-bonding.

Supporting Information. Error estimation for atomic coordinates. This material is available free of charge via the Internet at http://pubs.acs.org.

BIOGRAPHICAL INFORMATION

Martin Egli is Professor in the Department of Biochemistry at the School of Medicine, Vanderbilt University, Nashville, Tennessee, USA.

FOOTNOTES

*E-mail: martin.egli@vanderbilt.edu.
The author declares no competing financial interest.

REFERENCES

1   Moran, S.; Ren, R. X.-F.; Rumney, S.; Kool, E. T. Difluorotoluene, a Nonpolar Isostere of Thymine, Codes Specifically and Efficiently for Adenine in DNA Replication. *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057.

2   Moran, S.; Ren, R. X.; Kool, E. T. A Thymidine Triphosphate Shape Analog Lacking Watson-Crick Pairing Ability is Replicated With High Sequence Selectivity. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506–10511.

3   Morales, J. C.; Kool, E. T. Efficient Replication Between Non-Hydrogen-Bonded Nucleoside Shape Analogs. *Nat. Struct. Biol.* **1998**, *5*, 950–954.

4   Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2985–2993.

5   Koshland, D. E., Jr. The Key-Lock Theory and the Induced Fit Theory. *Angew. Chem., Int. Ed.* **1994**, *33*, 2375–2378.

6   Kool, E. T. Hydrogen Bonding, Base Stacking, and Steric Effects in DNA Replication. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 1–22.

7   Kool, E. T. Active Site Tightness and Substrate Fit in DNA Replication. *Annu. Rev. Biochem.* **2002**, *71*, 191–219.

8   Sintim, H. O.; Kool., E. T. The Difluorotoluene Debate - A Decade Later. *Chem. Commun.* **2006**, *35*, 3665–3675.

9   Lee, I.; Berdis, A. J. Non-Natural Nucleotides as Probes for the Mechanism and Fidelity of DNA Polymerases. *Biochim. Biophys. Acta* **2010**, *1804*, 1064–1080.

10 Howard, J. A. K.; Hoy, V. J.; O'Hagan, D.; Smith, G. T. How Good is Fluorine as a Hydrogen Bond Acceptor? *Tetrahedron* **1996**, *52*, 12613–12622.

11 Dunitz, J. D. Organic Fluorine: Odd Man Out. *ChemBioChem* **2004**, *5*, 614–621.

12 Rybalova, T. V.; Bagryanskaya, I. Y. C-F$\cdots\pi$, F$\cdots$H, and F$\cdots$F Intermolecular Interactions and F-Aggregation: Role in Crystal Engineering of Fluoroorganic Compounds. *J. Struct. Chem.* **2009**, *50*, 741–753.

13 Zhou, P.; Zou, J.; Tian, F.; Shang, Z. Fluorine Bonding — How Does It Work in Protein-Ligand Interactions? *J. Chem. Inf. Model.* **2009**, *49*, 2344–2355.

14 Chopra, D.; Guru Row, T. N. Role of Organic Fluorine in Crystal Engineering. *CrystEngComm* **2011**, *13*, 2175–2186.

15 Sun, Z.; McLaughlin, L. W. Probing the Nature of Three-Centered Hydrogen Bonds in Minor-Groove Ligand-DNA Interactions: The Contribution of Fluorine Hydrogen Bonds to Complex Stability. *J. Am. Chem. Soc.* **2007**, *129*, 12531–12536.

16 Schweitzer, B. A.; Kool, E. T. Hydrophobic, Non-Hydrogen-Bonding Bases and Base Pairs in DNA. *J. Am. Chem. Soc.* **1995**, *117*, 1863–1872.

17 Koller, A. N.; Bozilovic, J.; Engels, J. W.; Gohlke, H. Aromatic N Versus Aromatic F: Bioisosterism Discovered in RNA Base Pairing Interactions Leads to a Novel Class of Universal Base Analogs. *Nucleic Acids. Res.* **2010**, *38*, 3133–3146.

18 Guckian, K; Schweitzer, B. A.; Ren, R. X.-F.; Sheils, C J.; Paris, P. L.; Tahmassebi, D. C.; Kool, E. T. Experimental Measurement of Aromatic Stacking in the Context of Duplex DNA. *J. Am. Chem. Soc.* **1996**, *118*, 8182–8183.

19 Guckian, K. M.; Krugh, T. R.; Kool, E. T. Solution Structure of a Duplex DNA Containing a Replicable Difluorotoluene-Adenine Pair. *Nat. Struct. Biol.* **1998**, *5*, 954–959.

20 Wang, X.; Houk, K. N. Difluorotoluene, a Thymine Isostere, Does Not Hydrogen Bond at All. *Chem. Commun.* **1998**, 2631–2632.

21 Kim, T.; Brieba, L. G.; Ellenberger, T.; Kool, E. T. Functional Evidence for a Small and Rigid Active Site in a High Fdelity DNA Polymerase: Probing T7 DNA Polymerase with Variably Sized Base Pairs. *J. Biol. Chem.* **2006**, *281*, 2289–2295.

22 Potapova, O.; Chan, C.; DeLucia, A. M.; Helquist, S. A.; Kool, E. T.; Grindley, N. D. F.; Joyce, C. M. DNA Polymerase Catalysis in the Absence of Watson-Crick Hydrogen Bonds: Analysis by Single-Turnover Kinetics. *Biochemistry* **2006**, *45*, 890–898.

23 Lee, H. R.; Helquist, S. A.; Kool, E. T.; Johnson, K. A. Importance of Hydrogen Bonding for Efficiency and Specificity of the Human Mitochondrial DNA Polymerase. *J. Biol. Chem.* **2008**, *283*, 14402–14410.

24 Mizukami, S.; Kim, T.; Helquist, S. A.; Kool, E. T. Varying DNA Base-Pair Size in Subangstrom Increments: Evidence for a Loose, Not Large, Active Site in Low-Fidelity Dpo4 Polymerase. *Biochemistry* **2006**, *45*, 2772–2778.

25 Irimia, A.; Eoff, R. L.; Pallan, P. S.; Guengerich, F. P.; Egli, M. Structure and Activity of Y-class DNA Polymerase Dpo4 from Sulfolobus solfataricus with Templates Containing the

26 Washington, M. T.; Helquist, S. A.; Kool, E. T.; Prakash, L.; Prakash, S. Requirement of Watson-Crick Hydrogen Bonding for DNA Synthesis by Yeast DNA Polymerase η. *Mol. Cell. Biol.* **2003**, *23*, 5107–5112.

27 Wolfe, W. T.; Washington, M. T.; Kool, E. T.; Spratt, T. E.; Helquist, S. A.; Prakash, L.; Prakash, S. Evidence for a Watson-Crick Hydrogen Bonding Requirement in DNA Synthesis by Human DNA Polymerase κ. *Mol. Cell. Biol.* **2005**, *25*, 7137–7143.

28 Hogg, M.; Wallace, S. S.; Doublié, S. Bumps in the Road: How Replicative DNA Polymerases See DNA Damage. *Curr. Opin. Struct. Biol.* **2005**, *15*, 86–93.

29 Egli, M. DNA-Cation Interactions: Quo Vadis? *Chem. Biol.* **2002**, *9*, 277–286.

30 Pallan, P. S.; Egli, M. The Pairing Geometry of the Hydrophobic Thymine Analog 2,4-Difluorotoluene in Duplex DNA as Analyzed by X-ray Crystallography. *J. Am. Chem. Soc.* **2009**, *131*, 12548–12549.

31 Pallan, P. S.; Egli, M. Insights into RNA/DNA Hybrid Recognition and Processing by RNase H from the Crystal Structure of a Non-specific Enzyme-dsDNA Complex. *Cell Cycle* **2008**, *7*, 2562–2569.

32 Tereshko, V.; Minasov, G.; Egli, M. The Dickerson—Drew B-DNA Dodecamer Revisited at Atomic Resolution. *J. Am. Chem. Soc.* **1999**, *121*, 470–471.

33 Xia, J.; Noronha, A.; Toudjarska, I.; Li, F.; Akinc, A.; Braich, R.; Rajeev, K. G.; Egli, M.; Manoharan, M. siRNAs with a Ribo-difluorotoluyl Nucleotide: Structure, RISC-mediated Recognition, and Silencing. *ACS Chem. Biol.* **2006**, *1*, 176–183.

34 Somoza, A.; Chelliserrykattil, J.; Kool, E. T. The Roles of Hydrogen Bonding and Sterics in RNA Interference. *Angew. Chem., Int. Ed.* **2006**, *45*, 4994–4997.

35 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. Crystal Structure, Stability and in Vitro RNAi Activity of Oligoribonucleotides Containing the Ribo-difluorotoluyl Nucleotide: Insights into Substrate Requirements by the Human RISC Ago2 Enzyme. *Nucleic Acids Res.* **2007**, *35*, 6424–6438.

36 Zacharias, M.; Engels, J. W. Influence of a Fluorobenzene Nucleobase Analogue on the Conformational Flexibility of RNA Studied by Molecular Dynamics Simulations. *Nucleic Acids Res.* **2004**, *32*, 6304–6311.

37 Pfaff, D. A.; Clarke, K. M.; Parr, T. A.; Cole, J. M.; Geierstanger, B. H.; Tahmassebi, D. C.; Dwyer, T. J. Solution Structure of a DNA Duplex Containing a Guanine—Difluorotoluene Pair: A Wobble Pair without Hydrogen Bonding? *J. Am. Chem. Soc.* **2008**, *130*, 4869–4878.

38 Mueller, U.; Schubel, H.; Sprinzl, M.; Heinemann, U. Crystal Structure of Acceptor Stem of tRNA[Ala] from Escherichia coli Shows Unique G.U Wobble Base Pair at 1.16 Å Resolution. *RNA* **1999**, *5*, 670–677.

39 Manoharan, M.; Akinc, A.; Pandey, R. K.; Qin, J.; Hadwiger, P.; John, M.; Mills, K.; Charisse, K.; Maier, M. A; Nechev, L.; Greene, E. M.; Pallan, P. S.; Rozners, E.; Rajeev, K. G.; Egli, M. Unique Gene-Silencing and Structural Properties of 2′-F Modified siRNAs. *Angew. Chem., Int. Ed.* **2011**, *50*, 2284–2288.

40 Pallan, P. S.; Greene, E.; Jicman, P.; Pandey, R.; Manoharan, M.; Rozners, E.; Egli, M. Unexpected Origins of the Enhanced Pairing Affinity of 2′-Fluoro-Modified RNA. *Nucleic Acids Res.* **2011**, *39*, 3482–3495.

41 Li, F.; Sarkhel, S.; Wilds, C. J.; Wawrzak, Z.; Prakash, T. P.; Manoharan, M.; Egli, M. 2′-Fluoroarabino- and Arabinonucleic Acid Show Different Conformations, Resulting in Deviating RNA Affinities and Processing of Their Heteroduplexes with RNA by RNase H. *Biochemistry* **2006**, *45*, 4141–4152.

42 Yahyaee Anzahaee, M.; Watts, J. K.; Alla, N. R.; Nicholson, A. W.; Damha, M. J. Energetically Important C—H$\cdots$F—C Pseudohydrogen Bonding in Water: Evidence and Application to Rational Design of Oligonucleotides with High Binding Affinity. *J. Am. Chem. Soc.* **2011**, *133*, 728–731.

43 Xia, S.; Eom, S. H.; Konigsberg, W. H.; Wang, J. Structural Basis for Differential Insertion Kinetics of dNMPs Opposite a Difluorotoluene Nucleotide Residue. *Biochemistry* **2012**, *51*, 1476–1485.

Hydrophobic Thymine Analog 2,4-Difluorotoluene. *J. Biol. Chem.* **2007**, *282*, 36421–36433.