

A thymidine triphosphate shape analog lacking Watson–Crick pairing ability is replicated with high sequence selectivity

(nonpolar isostere/hydrogen bonding/base stacking/replication fidelity/Klenow fragment)

SEAN MORAN*, REX X.-F. REN*, AND ERIC T. KOOL*†‡

Departments of *Chemistry and †Biochemistry and Biophysics, University of Rochester, Rochester, NY 14627

Edited by Thomas C. Bruice, University of California, Santa Barbara, CA, and approved July 11, 1997 (received for review May 12, 1997)

ABSTRACT Compound **1** (**F**), a nonpolar nucleoside analog that is isosteric with thymidine, has been proposed as a probe for the importance of hydrogen bonds in biological systems. Consistent with its lack of strong H-bond donors or acceptors, **F** is shown here by thermal denaturation studies to pair very poorly and with no significant selectivity among natural bases in DNA oligonucleotides. We report the synthesis of the 5'-triphosphate derivative of **1** and the study of its ability to be inserted into replicating DNA strands by the Klenow fragment (KF, *exo*⁻ mutant) of *Escherichia coli* DNA polymerase I. We find that this nucleotide derivative (dFTP) is a surprisingly good substrate for KF; steady-state measurements indicate it is inserted into a template opposite adenine with efficiency (V_{\max}/K_m) only 40-fold lower than dTTP. Moreover, it is inserted opposite A (relative to C, G, or T) with selectivity nearly as high as that observed for dTTP. Elongation of the strand past **F** in an F–A pair is associated with a brief pause, whereas that beyond A in the inverted A–F pair is not. Combined with data from studies with **F** in the template strand, the results show that KF can efficiently replicate a base pair (A–F/F–A) that is inherently very unstable, and the replication occurs with very high fidelity despite a lack of inherent base-pairing selectivity. The results suggest that hydrogen bonds may be less important in the fidelity of replication than commonly believed and that nucleotide/template shape complementarity may play a more important role than previously believed.

DNA polymerases are well documented to make very few errors in the initial strand synthesis (1–3). For example, DNA polymerase I (pol I) from *Escherichia coli* [and its large (Klenow) fragment], one of the most extensively studied polymerases, is known to make an error in nucleotide insertion only once in 10^3 – 10^5 bases (4, 5). To begin to understand the origins of this selectivity it has been important to study the kinetics, thermodynamics, and structure of duplex DNA, of polymerase enzymes, and of their complexes in well defined systems. The insertion of a nucleotide at the end of a primer in a primer–template duplex occurs with at least three separate kinetic steps, and these have been characterized in detail recently for pol I (KF) and several other polymerases (4–7). Also central to the study of DNA polymerase fidelity has been the use of modified nucleoside analogs, primarily as altered bases in template strands, but in a number of cases also as modified nucleoside triphosphates. These compounds have been useful in examination of the importance of specific structural features and interactions in the efficiency and fidelity of enzymatic processing (1). The combination of detailed kinetic studies and the evaluation of the processing of nucleoside

analogues have led to the currently held models for DNA polymerase fidelity (1–10).

Perhaps the least understood aspect of polymerase fidelity is the question of which physical noncovalent interactions between nucleoside triphosphate, enzyme, and template are the strongest contributors to the choice of the correct nucleotide for insertion into the growing strand. It is clear that hydrogen bonding, solvation, base stacking, and steric (size and shape) factors all are conceivably important in the selection process. Of these, Watson–Crick hydrogen bonding has been cited most consistently as the source of specificity, and this belief exists at the level of teaching paradigms. For example, in their molecular biology textbook (11), Watson *et al.* argue that the specificity of hydrogen bonds alone may explain the observed fidelity. Similarly, Stryer's biochemistry text (12) states that "the likelihood of binding and . . . making a phosphodiester bond is very low unless the incoming nucleotide forms a Watson–Crick base pair with the opposing nucleotide on the template."

However, although DNA base analogs have proven useful in the study of polymerase fidelity, the analogs studied to date make it difficult to distinguish between hydrogen bonding and steric factors in nucleotide selection. The act of blocking or removing a hydrogen-bonding group by standard methods leads to significant changes in size and shape as well. It is usually found that such analogs are very poor polymerase substrates, and although this may be attributable to the loss of hydrogen bonds, it may easily be due to the size and shape differences instead.

To address this issue we have proposed four DNA base analogs that completely lack the polar Watson–Crick hydrogen-bonding groups of the natural bases but that retain their size and shape as closely as possible (13). For example, a nucleoside derived from difluorotoluene [compound **1** (**F**)] is virtually identical in size, shape, and conformation to thymidine (13, 42). Therefore, this compound seems ideal to test the relative importance of hydrogen bonds and steric effects in DNA synthesis, because it forms hydrogen bonds weakly, if at all, in water (15).

In an initial study we placed this compound into synthetic DNA templates and examined whether the Klenow fragment (KF) enzyme could insert natural nucleotides opposite it (17). Fidelity models purely dependent on hydrogen bonds would predict that the compound would serve both as a very poor substrate and with little or no selectivity. Surprisingly, however, the results showed very efficient DNA synthesis on this template and with high fidelity for insertion of adenine across from the difluorotoluene. These initial results led us to propose that hydrogen bonds alone may be less important in DNA synthesis than previously believed. More studies with the difluorotoluene analog were clearly warranted, in part because the known "A-rule" results in selective dATP insertion even in the absence of a template base (18–20).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0027-8424/97/9410506-6\$2.00/0 PNAS is available online at <http://www.pnas.org>.

This paper was submitted directly (Track II) to the *Proceedings* office. Abbreviations: KF, Klenow fragment; dFTP, difluorotoluene deoxynucleoside 5'-triphosphate.

A commentary on this article begins on page 10493.

‡e-mail: etk@etk.chem.rochester.edu.

We now report the synthesis of the nucleoside triphosphate derivative of this compound and its properties in DNA synthesis. Also described are the base-pairing properties of F in synthetic duplexes. The results provide further evidence that strong Watson-Crick hydrogen bonds are not required for highly specific DNA synthesis.

EXPERIMENTAL SECTION

Synthesis of Difluorotoluene Deoxynucleoside 5'-Triphosphate (dFTP). Difluorotoluene nucleoside (33 mg, 0.15 mmol) was dissolved in 0.75 ml trimethylphosphate, and the solution was cooled to 0°C. Proton Sponge [48 mg 1,8-bis(dimethylamino)naphthalene, (Aldrich)] and phosphorous oxychloride (16 μ L) were added (21), and the solution was stirred for 2 hr at 0°C. Tributylamine (0.22 ml) and tributylammonium pyrophosphate (130 mg in 1.7 ml dimethylformamide) were added (22), and the solution was stirred for 1 min before adding 15 ml of 1 M triethylammonium bicarbonate (pH 7.5) to quench the reaction. After stirring 20 min at room temperature, the reaction mixture was concentrated to 1–2 ml by lyophilization. The triphosphate was purified by anion exchange on a DEAE-cellulose column at 4°C with a 0.1–1.0 M gradient of triethylammonium bicarbonate (pH 7.5). The appropriate fractions were converted to the sodium salt of the triphosphate (23), and the concentration was determined using extinction coefficient (1,200 M⁻¹ cm⁻¹) for the nucleoside (15). Phosphorous NMR was taken in D₂O with 50 mM Tris (pH 7.5) and 2 mM EDTA; an external phosphoric acid standard was used. The phosphorous NMR spectrum of the purified triphosphate matched a published spectrum of ATP (24): doublet at 11 ppm (α -phosphate), triplet at 22 ppm (β -phosphate), and pH-sensitive doublet between 5 and 10 ppm (γ -phosphate).

Oligodeoxynucleotide Synthesis. DNA oligonucleotides were synthesized on an Applied Biosystems 392 synthesizer using standard β -cyanoethylphosphoramidite chemistry and were purified and quantitated as described previously (15).

Qualitative Primer Extension Reactions. Primers (23 nucleotides for single insertions, 18 for running start; see Fig. 1B) were labeled with [³²P]ATP and T4 polynucleotide kinase and purified using QIAquick anion-exchange spin columns (Qia-

gen). Labeled primer (final concentration, 20–25 nM) was mixed with buffer (10 mM Tris (pH 7.5), 5 mM MgCl₂, 7.5 mM DTT) and template (final concentration, 200 nM) and incubated for 5 min at room temperature; unlabeled primer was then added to give a final total concentration of primer-template of approximately 200 nM. After incubating another 5 min at room temperature, Klenow fragment (exo⁻) (New England Biolabs) pretreated with inorganic pyrophosphatase (see below) was added to a final concentration of 200 nM. Following a 2-min incubation at 37°C, triphosphates were added to a final concentration of 2 or 20 μ M. Reactions were run for 2-, 15-, or 60-min intervals and quenched with 0.8 volumes of loading buffer I [80% formamide, 1 \times TBE (89 mM Tris/borate [pH 8.3], 2 mM EDTA), 0.05% xylene cyanol, and bromophenol blue]. Reaction products were separated from unreacted primer by electrophoresis on 20% denaturing polyacrylamide gels and visualized by autoradiography.

The commercial polymerase showed evidence of contaminating pyrophosphate in control reactions. By pretreating the polymerase enzyme with approximately 0.0002 units of inorganic pyrophosphatase per unit of KF for 10 min at room temperature before adding the primer-template duplex, interfering pyrophosphorolysis reactions were prevented.

Steady-State Kinetics. Steady-state kinetics for standing-start single-nucleotide insertions were carried out as described (25). Briefly, insertion reactions were carried out in 10- μ l volumes in the presence of 50 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 1 mM DTT, and 0.1 mg/ml BSA. Primer-template duplexes were annealed by mixing in the buffer, heating to 90°C, and slow-cooling to 37°C. A 4 \times concentrated stock of duplex was mixed with 4 \times KF polymerase (exo⁻ mutant, Amersham-United States Biochemical) and incubated 2 min at 37°C. Reactions were initiated by adding the DNA-enzyme mixture to an equal volume of a 2 \times dNTP stock solution. The DNA (duplex) concentration was 5 μ M except where noted. Enzyme concentration and reaction time were adjusted in different dNTP reactions to give 1–20% incorporation in time periods \leq 20 min. Extents of reaction were determined by running quenched reaction samples on a 15% denaturing polyacrylamide gel to separate unreacted primer from insertion products; relative velocities were calculated as extent of reaction divided by reaction time and normalized for the varying enzyme concentrations used.

Steady-state kinetics for standing-start single-nucleotide insertions were carried out as described previously (17). Briefly, insertion reactions were carried out in the following buffers: 50 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 1 mM DTT, 0.1 μ g/ml BSA for Klenow fragment (exo⁻ mutant) obtained from Amersham-United States Biochemical, or 10 mM Tris-HCl (pH 7.4), 5 mM MgCl₂, 7.5 mM DTT (BSA added to 0.1 μ g/ml) for Klenow fragment (exo⁻) from New England Biolabs. Primer-template duplexes were annealed by mixing in buffer, heating to 90°C, and slow-cooling to 37°C. A 4 \times concentrated stock solution of duplex was mixed with 4 \times KF polymerase for 2 min at 37°C, and the reaction was initiated by adding a 2 \times solution of the appropriate dNTP. Quenching of the reaction was effected by adding 2–3 volumes of loading buffer II (95% formamide, 20 mM EDTA, 0.05% xylene cyanol and bromophenol blue). The final DNA (duplex) concentration was either 0.75 or 5 μ M. Amount of polymerase used and reaction time were adjusted to give 1–20% extent of reaction in reasonable time intervals. (The time for insertion of F across from G had to be substantially increased because this reaction is very inefficient.) The following ranges of enzyme concentrations and times were used (N \rightarrow M denotes dNTP inserted across from base M in the template strand): T \rightarrow A and F \rightarrow A, 7–33 nM, 0.5–3 min; T \rightarrow C, 13–67 nM, 5 and 13 min; F \rightarrow C, 130 and 200 nM, 10 and 36 min.; T \rightarrow G, 13–33 nM, 1.5–4 min; F \rightarrow G, 200 and 400 nM, 10 and 107 min; T \rightarrow T, 67 and 130 nM, 5–27 min; F \rightarrow T, 67–200 nM, 10 and 27 min; F \rightarrow abasic, 130 nM, 20 min. Extents of reaction were determined by running quenched reaction samples on 15 or 20% denaturing

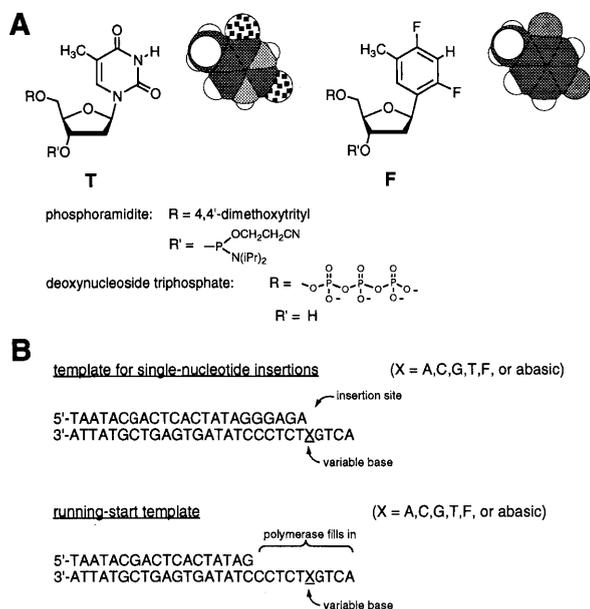


FIG. 1. The nucleotide structures and DNA sequences in this study. (A) Chemical structures of nucleosides T and analog F alongside space-filling models of the "nucleobases" of the two. (B) Sequences of template-primer duplexes used in steady-state studies. The standing-start substrate utilizes a 23-nt primer, whereas the running-start substrate utilizes an 18-nt primer.

polyacrylamide gels to separate unreacted primer from insertion products; relative velocities were calculated as extent of reaction divided by reaction time and normalized for the varying enzyme and duplex concentrations used.

Thermal Denaturation Experiments. Solutions for the thermal denaturation studies contained a 1:1 molar ratio of pyrimidine oligonucleotide and complementary 12-nt purine oligomer (1.5 μ M each). Solutions were buffered with 10 mM Na-PIPES [1,4-piperazine-bis(ethanesulfonate), Sigma], pH 7.0/100 mM NaCl/10 mM MgCl₂. The thermal denaturations were carried out as described previously (15).

RESULTS

Base Pairing Properties of Nonpolar Isostere F. We first utilized isostere F (Fig. 1) to examine the importance of canonical Watson–Crick hydrogen bonds in the stability and specificity of base pairing. We synthesized 12-mer oligodeoxynucleotides containing either thymine or difluorotoluene nucleosides at a central position where pairing selectivity of bases is at a maximum. We examined duplex stabilities when T or F is paired with complementary or mismatched partners to determine (i) the stability of F–A pairs in the context of duplex DNA and (ii) the inherent selectivity of pairing of F for A relative to C, G, and T. These were studied by UV-monitored thermal denaturation experiments at pH 7.0.

The results are presented in Table 1. The data show that the fully complementary, naturally substituted duplex is the most thermally and thermodynamically stable one of the series, with a T_m of 39°C and $-\Delta G_{25}^\circ$ of 12.3 kcal/mol. Placement of mismatched bases against a central thymine results in a lowering of T_m by 9–13°C and of stability (as measured by curve fitting) by 3.0–3.6 kcal/mol. By comparison, the complementary duplex in which the thymine is replaced by one difluorotoluene is considerably less stable, with T_m lowered by 12°C and stability decreased by 3.6 kcal/mol. This is quite similar in stability to the naturally mismatched duplexes (T paired with C, G, T). All four duplexes containing F fall into a narrow range of stability, with T_m values within 5.4°C of one another and free energies within 1.7 kcal/mol. Thus, the results show that F, although the same size and shape as thymine, is quite destabilizing when paired with adenine in this sequence and show, within experimental error, no thermodynamic selectivity for adenine over other partners. This finding is

Table 1. Free energies [$-\Delta G_{25}^\circ$ (kcal)] and melting temperatures [T_m (°C)] for duplexes containing a variable T–X or F–X base pair

Duplex	T_m (°C)*	$-\Delta G_{25}^\circ$ (kcal)
5'–CTTTTC \square TTTCTT 3'–GAAAAG \square AAGAA	39.4	12.3
5'–CTTTTC \square TTTCTT 3'–GAAAAG \square CAAGAA	26.4	8.7
5'–CTTTTC \square TTTCTT 3'–GAAAAG \square GAAGAA	30.7	9.3
5'–CTTTTC \square TTTCTT 3'–GAAAAG \square TAAGAA	27.1	8.9
5'–CTTTTC \square TTTCTT 3'–GAAAAG \square AAGAA	27.4	8.7
5'–CTTTTC \square TTTCTT 3'–GAAAAG \square CAAGAA	22.0	7.7
5'–CTTTTC \square TTTCTT 3'–GAAAAG \square GAAGAA	25.2	8.7
5'–CTTTTC \square TTTCTT 3'–GAAAAG \square TAAGAA	26.5	9.4

*Error in T_m values is estimated at $\pm 0.5^\circ\text{C}$, and in ΔG° values, $\pm 10\%$.

also consistent with previous studies of other nonpolar DNA base analogs (15, 26, 27).

Enzyme Substrate Properties of dFTP. To examine the importance of conventional Watson–Crick hydrogen bonding on the insertion of nucleotides by DNA polymerase, we prepared the nucleoside triphosphate derivative of the difluorotoluene nucleoside (Fig. 1) and studied its behavior with the Klenow fragment (exo[−] mutant) of DNA polymerase I from *E. coli*. Qualitative studies were carried out using 28-mer templates with 18-mer or 23-mer primers (Fig. 1B). The templates contain a variable base at the position 24 nucleotides downstream of the primer 5' end; thus, the 18-mer primer allows “running start” elongation up to and beyond the template variable site, whereas the 23-mer primer allows “standing start” single-nucleotide insertion opposite the variable site.

A single-nucleotide insertion survey was carried out for templates containing all four natural bases as well as F and using all four natural dNTPs as well as dFTP (Fig. 2). It was carried out under nonforcing conditions (2 μ M dNTP, 200 nM primer–template duplex, 200-nM enzyme, 2-min reaction) so that inefficiently processed base pairs, such as base mismatches, were not formed. The results show that the four natural base pairs are formed (with essentially complete elongation of the primer) as expected. Interestingly, three other base pair insertions are also observed: dFTP opposite A, dATP opposite F, and dFTP opposite F. Thus, the polymerase is able to efficiently process three new DNA base pairs (the base listed first is inserted as a dNTP opposite the template base, listed second): F \rightarrow A, A \rightarrow F, and F \rightarrow F. The insertion of A opposite F was recently reported (17), but the F \rightarrow A and F \rightarrow F pairs were not known until the dFTP analog was available (this work).

Efficient nucleotide insertion does not guarantee successful replication, since mispairs and lesions can often cause the polymerase to stall after insertion (preventing efficient formation of the next base pair) (28–30). Therefore, we tested the ability of the KF polymerase to synthesize a strand continuing beyond an F–A pair by using the 18-mer primer and template with an A at the variable site, six nucleotides downstream of the primer 3' end (Fig. 1B). The dNTP mixture containing dATP, dCTP, dGTP, and dFTP was used (20 μ M each). Once again, nonforcing conditions were used (20 μ M dNTP, 200 nM primer–template duplex, 200 nM enzyme).

The results are shown in Fig. 3. The autoradiogram shows complete elongation of the primers to the variable position at the earliest (2 min) time in two templates. An abasic template shows a long-lived pause (≥ 60 min) after insertion opposite the abasic site. By contrast, an adenine-containing template shows only a short-lived pause, which is visible only at 2 min and is fully elongated at 15 min. Experiments in which dFTP is omitted from the mixture result in a long-lived pause prior to the adenine site (data not shown), adding confirmation that dFTP is inserted there.

Steady-State Kinetics for Insertion of dFTP. To compare efficiencies of enzymatic formation of natural base pairs with those involving dFTP, we carried out a series of kinetics studies using steady-state methods (25, 31). We used the standing-start primer–template duplex, as above, and studied single-nucleotide insertions by varying dNTP concentrations under single-turnover conditions. We studied the insertion of dFTP opposite natural bases as well as difluorotoluene and the abasic site, and for comparison we studied the insertion of dTTP opposite the natural bases and difluorotoluene. The kinetic parameters derived from the data are given in Table 2.

The kinetic data fit the linear Hanes–Woolf form well, giving R values of 0.99 or better (data not shown), and the results confirm the qualitative observations seen earlier. For insertion of dTTP, we find that the template containing A is processed from 7×10^3 to 9×10^4 times more efficiently (as V_{\max}/K_m) than the mismatched (C, G, T) templates, as observed previously. We find that dFTP insertion opposite A is surprisingly efficient, being only

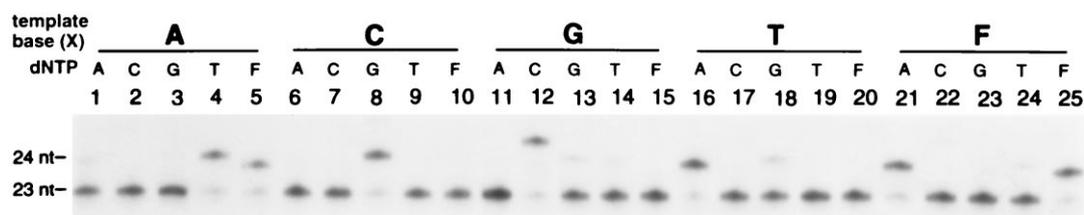


FIG. 2. Autoradiogram showing a survey of standing-start single-nucleotide insertions by the KF (*exo*⁻) enzyme, including all possible cases of analog F in the template and as a dNTP. The data were taken using 2 μ M dNTP, and the reactions were stopped after 2 min.

40-fold less efficient than dTTP with the same template. Even more interesting is the finding that dFTP is inserted with high selectivity opposite A relative to C, T, and G. Here, the template containing A is processed from 3×10^3 to 1×10^4 times more efficiently than the mismatched (C, G, T) templates, and so the selectivity for the correct template is only ≈ 2 - to 7-fold lower than that with the natural nucleotide dTTP.

Also examined by this approach were nonnatural pyrimidine isostere-pyrimidine pairs involving analog F. The nucleotide dTTP is found to be inserted opposite F with low efficiency about equal to, or perhaps slightly better than, the misinsertion opposite G, which is again consistent with the earlier qualitative studies. Those studies did find that dFTP is inserted opposite F, however, and the kinetic data confirm this, giving a V_{\max}/K_m value approximately 40-fold higher than insertion of T opposite F and two-thirds that of dFTP insertion opposite A.

The efficiency (V_{\max}/K_m) data are also plotted as a histogram in Fig. 4 to allow ready comparison. Here we compare the relative efficiencies and selectivities using dTTP and the isosteric analog dFTP. Again, the shapes of the histograms, and thus the similarity in overall selectivity, is striking. The only easily distinguishable difference appears to be the efficiency of insertion opposite G, which is the most efficient of misinsertions with dTTP but the least efficient of those with dFTP. This appears quite similar to a histogram of earlier data with the reverse case, in which T or F

are situated in the template and varied nucleotides are inserted opposite them (17).

DISCUSSION

The Importance of Hydrogen Bonds in DNA Pairing. The data clearly add support to the conclusion that hydrogen-bonding complementarity plays an important role in stabilizing base pairs in the absence of polymerase enzymes (32). The thermal denaturation data show that the difluorotoluene base analog is considerably destabilizing when paired with adenine internally in a duplex. We attribute this ≈ 3.6 -kcal destabilization (relative to a T-A pair at this position) to loss of the hydrogen bonds in the pair and, probably more importantly, to the cost of desolvation of adenine's hydrogen-bonding groups when placed in pairing geometry (15, 33), although whether the F-A pair actually adopts a Watson-Crick-like geometry is not yet known. Because difluorotoluene base stacks considerably more strongly than thymine (34), the observed destabilization is probably an underestimate of the cost of pairing with adenine. Finally, with the natural bases it is clear that geometry also plays an important role in base pair stabilization, because base mismatches, like correctly matched base pairs, very often have two H-bonds but still are considerably destabilizing (35, 36).

In addition to forming unstable pairs, the nonpolar thymine analog displays little or no inherent selectivity for pairing with any of the four natural bases, which is consistent with previous studies of nonpolar DNA base analogs (15). We again attribute this result to the lack of strong hydrogen-bonding groups and to the cost of desolvation of the natural bases paired with it. It is clear that shape complementarity alone does not give the needed pairing selectivity for DNA, although for the natural bases a combination of hydrogen-bonding complementarity and the ability to adopt Watson-Crick geometry with those hydrogen bonds are both important for selectivity of pairing (35, 36). These data are valuable for two reasons. First, this is the first case of such pairing measurements in which steric effects can be effectively ruled out. Second, in the light of our finding that such a nonpolar nucleotide analog can be replicated, it is important to evaluate whether efficiency and selectivity of DNA synthesis arise from inherent properties of the base itself or whether they come chiefly from the enzyme.

Enzymatic Processing of an Unstable Base Pair. The steady-state method allows a simple comparison of the aggregate of kinetic steps leading to extension of a primer by one nucleotide (25). The individual kinetic steps involved in this process have been elucidated recently for the Klenow enzyme, and study of these individual steps requires time-resolved methods such as the following of pre-steady-state burst kinetics (6). The present data therefore are most useful in addressing the physical interactions that contribute to the overall process in the steady state.

Although our pairing data add support to the conclusion that hydrogen bonds are important for stabilization of DNA after it is synthesized, our polymerase studies indicate that these bonds may not be as directly important during the synthesis of the DNA. Our results demonstrate clearly that, at least under these conditions, difluorotoluene (F) is handled by the KF polymerase almost as if it were thymine. Our previous results showed that F acts in a template strand to encode

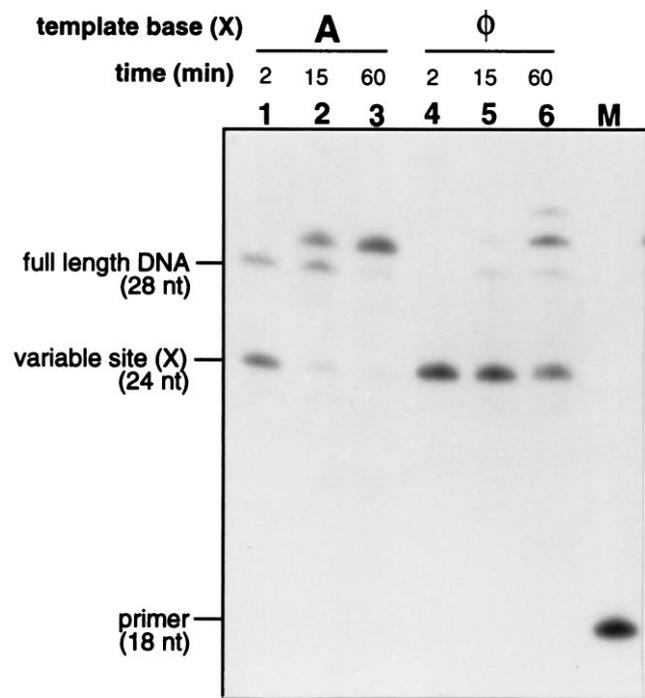


FIG. 3. Autoradiogram showing running-start primer elongations up to and beyond adenine with comparison with the known pause opposite an abasic nucleotide (ϕ). Reactions were performed with 20 μ M of each dNTP; for the template containing A, the dNTP mixture was dATP + dCTP + dGTP + dFTP, and for the abasic template, dATP + dCTP + dGTP + dTTP.

Table 2. Accuracy of template base selection for a given nucleoside triphosphate

Nucleoside triphosphate	Template base (X)	K_m , μM	V_{\max} , % min^{-1}	Efficiency, V_{\max}/K_m	Accuracy*
dTTP	A	2.5 (0.8)	43 (4)	1.7×10^7	1
	C	2,200 (800)	0.39 (0.07)	1.8×10^2	1.1×10^{-5}
	G	580 (120)	1.5 (0.1)	2.6×10^3	1.5×10^{-4}
	T	180 (20)	0.079 (0.002)	4.4×10^2	2.6×10^{-5}
	F	170 (30)	1.2 (0.4)	7.1×10^3	4.2×10^{-4}
dFTP	A	95 (110) [†]	15 (10) [†]	$4.2 \times 10^{5\dagger}$	1
	C	370 (90)	0.038 (0.002)	1.0×10^2	2.4×10^{-4}
	G	190 (40)	0.0056 (0.0003)	2.9×10^1	6.9×10^{-5}
	T	440 (140)	0.064 (0.005)	1.5×10^2	3.6×10^{-4}
	F	53 (4)	16 (1)	3.0×10^5	7.1×10^{-1}
	ϕ	79 (20)	0.039 (0.002)	4.9×10^2	1.2×10^{-3}

Steady-state kinetic parameters are for incorporation of single nucleotides into a template–primer duplex containing variable base X by the KF (exo⁻) polymerase. Conditions: 5 μM template–primer duplex, 13–200 nM enzyme, 50 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 1 mM dithiothreitol, and 100 $\mu\text{g}/\text{ml}$ bovine serum albumin, incubated for 1–20 min at 37°C in a reaction volume of 10 μl . Error limits are given in parentheses.

*Efficiency relative to dTTP insertion opposite A (entries 1–5) or dFTP insertion opposite A (entries 6–11).

[†]Value averaged from four data sets.

insertion of adenine with efficiency only 4-fold lower than if thymine is in the template (17). One might be tempted to explain this by citing the “A-rule,” wherein adenine is preferentially inserted opposite abasic sites (18–20). However, the data also showed that adenine was inserted with efficiency nearly two orders of magnitude higher opposite F than opposite an abasic site, and, moreover, there was no observable pause after the A–F pair, whereas there was a long-lived pause after the A-abasic site “pair.” These results argued against a simple A-rule explanation. Indeed, by using the steady-state method, it is difficult to observe any measurable difference between thymine and difluorotoluene in a template strand.

The present data, which differ from the previous study in that the analog F is used as a nucleoside triphosphate rather than as a template base, are of value because they allow a second test of this explanation. The new data clearly support those previous findings, because reversing the base pair has now been shown to still allow relatively efficient replication. The A-rule cannot be cited here, because A is not being inserted at the site in question.

Therefore, we must conclude that although difluorotoluene is inherently quite destabilizing to DNA, it still satisfies most of the physical criteria by which this polymerase accepts a base in a template or as a nucleoside triphosphate. Not only is stable base pairing not required, but in addition, canonical Watson–Crick hydrogen bonding cannot be an absolute requirement for efficient replication. Here one may be tempted to cite the possible formation of weak H-bonds of the type C–F··H–N or C–H··N

as sources of stabilization, but recent calculations (37), measurement of H-bonding ability of difluorotoluene in chloroform (15), and the present base-pairing data all point out that such effects must be quite small. The difluorotoluene shows no experimentally measurable tendency to undergo hydrogen bonding even in situations (such as location opposite adenine in DNA) that would strongly favor it. Indeed, the interaction of F with A in water is thermodynamically strongly repulsive because of the strong preference of A to be solvated by water.

It is also clear that the energetics of the F–A pair (as measured internally in the duplex) are not felt until after the enzymatic transition state for formation of the base pair, or the efficiency of processing the pair would be similar to a base mismatch (it is not; see Fig. 2). It has been pointed out that the pairing at the end of a strand is the most germane quantity for DNA polymerases (8), and our previous results with difluorotoluene show that it is at least as stable as thymine when situated at the end of a duplex (34). The present data do show, however, that dFTP insertion is less efficient than dTTP insertion by ≈ 40 -fold. Two possible sources of this difference are the loss of an H-bond between the polymerase and the base moiety, and/or the cost of partial desolvation of the template adenine at the transition state. Because the steady-state method used here does not address which step is rate-limiting in the observed insertion, pre-steady-state studies of dFTP replication would be warranted to investigate the mechanistic differences in processing of the difluorotoluene and thymine nucleotides in more detail.

High Replication Selectivity in the Absence of Canonical Hydrogen Bonds. Probably more surprising than the finding of efficient replication of analog F is the finding of high specificity for enzymatic pairing of difluorotoluene with adenine. Our pairing studies show that F has little or no inherent pairing selectivity for the natural bases even in the center of a standard duplex. Nevertheless, in the template it is replicated with fidelity very nearly as high as thymine. Similarly, as the dNTP analog, F is inserted opposite A with selectivity that is less than dTTP by only a small amount. Thus, the enzyme cannot be simply enforcing the base’s own preferences but must be contributing actively to pairing specificity.

A Replication Model Involving Shape Complementarity. From the results, we are led to conclude that canonical Watson–Crick hydrogen bonds are not required for high fidelity of replication. Although hydrogen bonds are clearly formed in the course of natural DNA replication and may well play an important role, we believe that the present results require some adjustment to beliefs that hydrogen bonds are the only force serving as the root of this high fidelity. The data argue that models that cite the number and

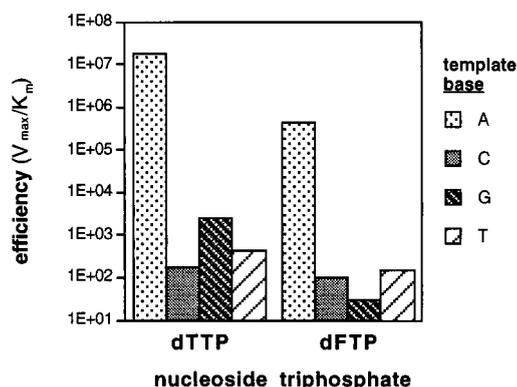


FIG. 4. Histogram showing accuracy of template selection for nucleotide insertion by the KF enzyme. Semilog plot for insertion efficiency of dFTP and dTTP opposite each of the four natural bases in a template.

strength of hydrogen bonds as the chief source of fidelity must be modified, because our nucleoside analog has none of the canonical Watson–Crick hydrogen-bonding groups, gives no measurable indication of forming hydrogen bonds, and yet is replicated with good efficiency and high fidelity. Therefore, we argue that shape complementarity is likely to be the most important criterion and hydrogen bonds might even be eliminated if the steric requirements are met.

Some have recognized that, in addition to complementary hydrogen bonding, matching of Watson–Crick geometry might be an important factor in replication fidelity (2, 3, 38). Petruska and Goodman and others (8, 39) have pointed out that one mechanism by which a DNA polymerase can enhance insertion specificity at the primer terminus is by forming an active site that closely fits the Watson–Crick geometry, thus restricting rotational freedom at the primer end and lowering the entropic cost of correct base pairing. Entropy–enthalpy compensation might increase favorable base pairing enthalpy at the transition state as well. Because the present difluorotoluene nucleoside analog, unlike previous analogs, has no significant steric differences from the natural nucleoside, it allows for a useful test of such a “goodness of fit” mechanism. Our results add support to this mechanism, because an analog isosteric with T is replicated with quite high fidelity. However, our results go further in suggesting that the H-bonding component of base pairing may not be the most directly important selectivity criterion and that shape exclusion may be more directly relevant. A close fit of the Watson–Crick geometry by the enzyme would not only lower the entropic cost of base pairing but would also result in a high enthalpic cost (because of steric clashes) for misinsertion of a nucleotide lacking the correct shape.

We surmise that this efficient replication of Watson–Crick-modified bases generally had not been observed previously because of the differences in shape between the bases and natural pyrimidines and purines. Steric (size and shape) effects are clearly of sufficient magnitude to explain the fidelity observed in DNA synthesis, especially in a geometrically constrained system defined by double-stranded DNA closely surrounded by the enzyme active site. Hence, very small differences in size and shape can result in large differences in reaction rates.

The present shape-exclusion model would appear to be consistent with results for known modified bases. For example, several bases covalently modified by alkylating agents have been studied; among the smallest of these lesions is the methylation at O⁴ of thymine. This lesion results in unstable duplex pairing and causes adenine to be inserted with efficiency that is hundreds of times lower than when it is opposite unmodified thymine (40, 41). The present model explains this by citing the sterically bulky methyl group as being too large to fit in the Watson–Crick geometry with adenine. Moreover, the misincorporation of guanine opposite this lesion can be explained as arising from the better shape complementarity with guanine, because the absence of the N-3 proton of T (as a result of methylation) makes room for the corresponding proton on G.

Nonpolar shape mimics such as the difluorotoluene nucleoside are likely to be generally useful in the study of polymerase mechanisms. Continuing studies will address the generality of these findings by examining other sequence contexts and other polymerase enzymes. It will also be of interest to examine related nonpolar DNA base analogs that vary in size and shape.

We thank the National Institutes of Health (GM52956) for support.

- Kornberg, A. & Baker, T. A., eds. (1992) *DNA Replication* (Freeman, New York), 2nd Ed., p. 113.
- Loeb, L. A. & Kunkel, T. A. (1982) *Annu. Rev. Biochem.* **52**, 429.
- Echols, H. & Goodman, M. F. (1991) *Annu. Rev. Biochem.* **60**, 477–511.
- Travaglini, E. C., Mildvan, A. S. & Loeb, L. A. (1975) *J. Biol. Chem.* **250**, 8647–8656.
- El Deiry, W. S., Downey, K. M. & So, A. G. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7378–7382.
- Kuchta, R. D., Mizrahi, V., Benkovic, P. A., Johnson, K. A. & Benkovic, S. J. (1987) *Biochemistry* **26**, 8410–8417.
- Kuchta, R. D., Benkovic, P. A. & Benkovic, S. J. (1988) *Biochemistry* **27**, 6716–6725.
- Petruska, J., Goodman, M. F., Boosalis, M. S., Sowers, L. C., Cheong, C. & Tinoco, I., Jr. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6252–6256.
- El Deiry, W. S., So, A. G. & Downey, K. M. (1988) *Biochemistry* **27**, 546–553.
- El Deiry, W. S., Downey, K. M. & So, A. G. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7378–7382.
- Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. & Weiner, A. M. (1987) *Molecular Biology of the Gene* (Benjamin/Cummings, Menlo Park, CA), 4th Ed., p. 283.
- Stryer, L. (1995) *Biochemistry* (Freeman, New York), 4th Ed., p. 89.
- Schweitzer, B. A. & Kool, E. T. (1994) *J. Org. Chem.* **59**, 7238–7242.
- Schweitzer, B. A. & Kool, E. T. (1995) *J. Org. Chem.* **60**, 8326.
- Schweitzer, B. A. & Kool, E. T. (1995) *J. Am. Chem. Soc.* **117**, 1863–1872.
- Schweitzer, B. A. & Kool, E. T. (1995) *J. Am. Chem. Soc.* **118**, 931.
- Moran, S., Ren, R. X.-F., Rumney, S. & Kool, E. T. (1997) *J. Am. Chem. Soc.* **119**, 2056–2057.
- Sagher, D. & Strauss, B. (1983) *Biochemistry* **22**, 4518–4526.
- Randall, S. K., Eritja, R., Kaplan, B. E., Petruska, J. & Goodman, M. F. (1987) *J. Biol. Chem.* **262**, 6864–6870.
- Takeshita, M., Chang, C.-N., Johnson, F., Will, S. & Grollman, A. P. (1987) *J. Biol. Chem.* **262**, 10171–10179.
- Kovacs, T. & Ötvös, L. (1988) *Tetrahedron Lett.* **29**, 4525–4528.
- Mishra, M. C. & Broom, A. D. (1991) *J. Chem. Soc. Chem. Commun.* 1276–1277.
- Hoard, D. E. & Ott, D. G. (1965) *J. Am. Chem. Soc.* **87**, 1785–1788.
- Jaffe, E. K. & Cohn, M. (1978) *Biochemistry* **17**, 652–657.
- Goodman, M. F., Creighton, S., Bloom, L. B. & Petruska, J. (1993) *Crit. Rev. Biochem. Mol. Biol.* **28**, 83–126.
- Millican, T. A., Mock, G. A., Chauncey, M. A., Patel, T. P., Eaton, M. A. W., Gunning, J., Cutbush, S. D., Neidle, S. & Mann, J. (1984) *Nucleic Acids Res.* **12**, 7435–7453.
- Nichols, R., Andrews, P. C., Zhang, P. & Bergstrom, D. E. (1994) *Nature (London)* **369**, 492–493.
- Mendelman, L. V., Petruska, J. & Goodman, M. F. (1990) *J. Biol. Chem.* **265**, 2338–2346.
- Dosanjh, M. K., Galeros, G., Goodman, M. F. & Singer, B. (1991) *Biochemistry* **30**, 11595–11599.
- Shibutani, S., Takeshita, M. & Grollman, A. P. (1991) *Nature (London)* **349**, 431–434.
- Boosalis, M. S., Petruska, J. & Goodman, M. F. (1987) *J. Biol. Chem.* **262**, 14689–14696.
- Kudritskaya, Z. G. & Danilov, V. I. (1976) *J. Theor. Biol.* **59**, 303–318.
- Alber, T., Dao-pin, S., Wilson, K., Wozniak, J. A., Cook, S. P. & Matthews, B. W. (1987) *Nature (London)* **330**, 41–46.
- Guckian, K., Schweitzer, B. A., Ren, R. X.-F., Sheils, C. J., Paris, P. L., Tahmassebi, D. C. & Kool, E. T. (1996) *J. Am. Chem. Soc.* **118**, 8182–8183.
- Brown, T., Hunter, W. N., Kneale, G. & Kennard, O. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2402–2406.
- Kouchakdjian, M., Li, B. F. L., Swan, P. F. & Patel, D. J. (1988) *J. Mol. Biol.* **202**, 139–155.
- Howard, J. A. K., Hoy, V. J., O’Hagan, D. & Smith, G. T. (1996) *Tetrahedron* **52**, 12613–12622.
- Joyce, C. M., Sun, X. C. & Grindley, N. D. F. (1992) *J. Biol. Chem.* **267**, 24485–24500.
- Petruska, J. & Goodman, M. F. (1995) *J. Biol. Chem.* **270**, 746–750.
- Singer, B., Spengler, S. J., Fraenkel-Conrat, H. & Kusmierek, J. T. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 28–32.
- Singer, B., Chavez, F., Spengler, S. J., Kusmierek, J. T., Mendelman, L. & Goodman, M. F. (1989) *Biochemistry* **28**, 1478–1483.
- Guckian, K. & Kool, E. T. (1997) *Angew. Chem. Int. Ed. Engl.*, in press.