Structural Factors That Determine Selectivity of a High Fidelity DNA Polymerase for Deoxy-, Dideoxy-, and Ribonucleotides

Weina Wang, Eugene Y. Wu, Homme W. Hellinga, and Lorena S. Beese*

From the Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710

Background: Nucleotide selection based on sugar moiety by DNA polymerases is essential for maintaining genomic integrity.

Results: Multiple polymerase conformations trap incorrect nucleotides.

Conclusion: In addition to simple steric blocks, non-cognate nucleotides are prevented from incorporating by ensembles of binding sites that stabilize substrate misalignment.

Significance: A structural basis has been revealed for ensembles of incorporation pathways previously defined by enzyme kinetics.

In addition to discriminating against base pair mismatches, DNA polymerases exhibit a high degree of selectivity for deoxyribonucleotides over ribo- or dideoxynucleotides. It has been proposed that a single active site residue (steric gate) blocks productive binding of nucleotides containing 2'-hydroxyls. Although this steric gate plays a role in sugar moiety discrimination, its interactions do not account fully for the observed behavior of mutants. Here we present 10 high resolution crystal structures and enzyme kinetic analyses of Bacillus DNA polymerase I large fragment variants complexed with deoxy-, ribo-, and dideoxynucleotides and a DNA substrate. Taken together, these data present a more nuanced and general mechanism for nucleotide discrimination in which ensembles of intermediate conformations in the active site trap non-cognate substrates. It is known that the active site O-helix transitions from an open state in the absence of nucleotide substrates to a ternary complex closed state in which the reactive groups are aligned for catalysis. Substrate misalignment in the closed state plays a fundamental part in preventing non-cognate nucleotide misincorporation. The structures presented here show that additional O-helix conformations intermediate between the open and closed state extremes create an ensemble of binding sites that trap and misalign non-cognate nucleotides. Water-mediated interactions, absent in the fully closed state, play an important role in formation of these binding sites and can be remodeled to accommodate different non-cognate substrates. This mechanism may extend also to base pair discrimination.

DNA polymerases impose exquisite specificity for correctly paired nucleotide base pair formation through filtering mechanisms that serially select against non-cognate pairings (1–6), thereby achieving the degree of fidelity required to maintain genomic integrity (7). Additionally, DNA polymerases select nucleotides on the basis of their sugar identity, discriminating between the deoxyribonucleotide (dNTP) and their corresponding, 10–2,000-fold more abundant ribonucleotide (rNTP) pools on the basis of the 2’ sugar hydroxyl present only in the latter (8–11). In addition to imposing the logic for the flow of genetic information (12), separating the synthesis of RNA and DNA also is necessary to prevent misincorporation of otherwise correctly paired ribonucleotides, which leads to genome instability (13). Ribonucleotide misincorporation possibly is the most common replication error in initial DNA synthesis (10) or subsequent gap filling (11), slowing down replication (10, 14) and increasing susceptibility to strand cleavage. DNA polymerases also recognize the 3’-OH that is absent in synthetic dideoxyribonucleotides (ddNTPs) that are used for DNA sequencing and some antiviral drugs (15). Understanding the mechanism by which DNA polymerases discriminate between different nucleotide pools based on the sugar moiety hydroxylation state is therefore of both fundamental and technological interest. Insights may aid in the development of therapeutics and synthetic RNA polymerases that enable RNA synthesis to be initiated using oligonucleotides rather than promoters (16, 17).

Although some specialized DNA polymerases are non-selective (e.g. human polymerase μ and human terminal DNA transferases), most DNA polymerases exhibit appreciable preference...
for dNTP over rNTP incorporation, with widely varying selectivity coefficients that range from \(10^{-6}\) to \(10^{6}\) (11). Mutagenesis studies of both high (14, 18–25) and low fidelity (26, 27) DNA polymerases have shown that mutation of a single residue in the active site significantly lowers the sugar-based selectivity (11). Molecular modeling suggested that this residue could interfere with binding of the rNTP 2'-OH and that reduction of its steric bulk accounts for the lowering of specificity in these mutants (14, 28). Structural studies of a high (23) and low fidelity (26, 27) DNA polymerases have shown that mutation of a single residue in the active site significantly lowers the sugar-based selectivity (29) DNA polymerase are consistent with this “steric gate” hypothesis.

Although the putative steric gate residue plays an important role in sugar specificity, the steric gate hypothesis does not account fully for the observed behavior of the mutants. First, in all cases, the decrease in sugar-based nucleotide selectivity is accompanied by a loss of dNTP incorporation rates (Fig. 1 and Table 1), indicating that the effects of the steric gate mutation on dNTP and rNTP incorporation are not independent. Second, even in the mutants, dNTP remains favored over rNTP on dNTP and rNTP incorporation are not independent. Specifically active closed state. Furthermore, we show that the E568A steric gate mutation results in rearrangements in the active site that can account for the consistent loss of dNTP incorporation.

DNA can be replicated in BF crystals, which has enabled many stages of the polymerase catalytic cycle to be captured crystallographically with a variety of substrates (31–34, 37–39). Conformational changes are key to the mechanism by which BF imposes a high degree of specificity. Briefly, dNTPs are incorporated at the insertion site in the “closed” conformational state of the DNA polymerase. In this form, the incoming base is paired with the template strand, and its α-phosphate is aligned with the 3'-OH terminus of the primer strand in the presence of Mg2+, which enters the active site together with the dNTP. In the absence of dNTPs, the polymerase adopts an “open” conformation. These two conformational states are determined primarily by the motion of the O helix in the presence of dNTPs, which swings from the “open” state (32, 40), through intermediates that in some instances can be captured crystallographically (38, 39), to a “closed” form (32, 40, 41). This motion combines bending of a hinge region at the O helix C terminus (residue Gly-711) with deformations within the helix (38). Base pair mismatches are selected against by a variety of filters that distinguish correctly from incorrectly paired nucleotides by their hydrogen bonding pattern or steric shape (33, 38, 39). Prior to incorporation, the α-phosphate of mismatched nucleotides or lesions misaligns with the primer terminus 3’-OH, thereby reducing or preventing the chemical incorporation.

### Table 1

**Summary of kinetic data of DNA polymerases from five families**

| Protein Family | Protein Name | N | WT | Steric gate mutant | Loss (dNTP)* | Gain (rNTP)* | ΔΔG° (dNTP)† | ΔΔG° (rNTP)† |
|----------------|--------------|---|----|-------------------|-------------|--------------|-------------|-------------|
| **Family A**   | *Escherichia coli* Klenow fragment (14) | C | 3400 | 4.3 | 34 | 23 | 2.1 | -1.8 |
|                | *Taq Pol I* (18) | dTTP/rUTP | 1,700,000 | 33 | 43 | 1200 | 2.2 | -4.2 |
|                | *Human Pol γ* (19) | A | 29,000 | 5.3 | 59 | 91 | 2.7 | -2.9 |
|                |               | C | 24,000 | 130 | 8.5 | 21 | 1.4 | -2.0 |
|                |               | C | 22,000 | 10 | 45 | 48 | 2.5 | -2.5 |
|                | *Bacillus* fragment (this study) | dTTP/rUTP | 77,000 | 222 | 76 | 4.3 | 2.7 | -0.90 |
| **Family B**   | *RB69* (20) | C | 24,000 | 170 | 46 | 3.0 | 2.2 | -0.64 |
|                | *φ 29* (21) | dTTP/rUTP | 4,400,000 | 2300 | 1.9 | >1000 | 0.35 | -3.8 |
| **Family X**   | *Human Pol λ* (22) | dTTP/rUTP | 3000 | 360 | 1.2 | 7.8 | 0.11 | -1.3 |
|                | *Human Pol β* (23) | C | 8200 | 670 | 2.3 | 5.5 | 0.51 | -1.1 |
| **Family Y**   | *Dpo4* (26) | G | 18,333 | 4 | 32 | 1367 | 2.1 | -4.4 |
|                |               | C | 5500 | 12 | 10 | 44 | 1.4 | -2.3 |
|                |               | A | 13,448 | 3 | 19 | 224 | 1.8 | -3.3 |
|                | *Dbh* (27) | dTTP/rUTP | 30,300 | 30 | 2.7 | 275 | 0.61 | -3.5 |
| **Reverse transcriptase** | *HIV-1 RT* (24) | G | 3400 | 3.7 | 28 | 33 | 1.95 | -2.1 |
|                |               | C | 7700 | 5 | 164 | 9.1 | 3.0 | -1.3 |
| **Moloney murine leukemia virus RT** (25) | dTTP/rUTP | 16,000 | 23 | 1.4 | 500 | 0.21 | 0.21 | -3.8 |

*ΔΔG° (dNTP) = ΔG° (WT) - ΔG° (Mutant)*

*ΔΔG° (rNTP) = ΔG° (rNTP) WT - ΔG° (rNTP) Mutant*
step (37–39). Interference with the motion of the O helix is an important component of these misalignment processes; incorrect dNTPs positioned at the insertion site form only partially closed O helix conformations (38, 39). Following their incorporation, mismatches or lesions introduce distortions of the template strand and alter the alignment of the 3’-OH to interfere with subsequent incorporation events (33–36). Here we report that in the sugar moiety discrimination process, O helix motions also are interfered with, resulting in non-productive conformations that differ from those observed for mismatches or lesions.

**EXPERIMENTAL PROCEDURES**

**Protein Preparation**

Our *Bacillus* fragment DNA polymerase I originated from a thermophilic bacterium that was isolated from a calcium carbonate mound of a neutral pH hot spring in Idaho (30). It was originally classified as a strain of *Bacillus steathermophilus* based on 16S rRNA sequence analysis (30). The most closely sequenced homolog is from *Geobacillus kaustophilus* with a single mutation, H823R (UniProtKB: Q5KWC1). D598A, F710Y/D598A, E658A/D598A, F710Y, and E658A mutants were constructed using QuikChange site-directed mutagenesis (Stratagene, La Jolla, CA). Wild-type and mutant proteins were expressed and purified as described (30). Proteins containing the D598A mutation were used for crystallization to obtain closed form ternary complexes 1–9 (Table 3) of BF polymerase, DNA duplex, and incoming nucleotide substrate (paired with complementary template base) positioned at the insertion site prior to chemistry (32). Wild-type protein was used to obtain the open form, binary complex 10 (Table 3) with ribonucleotide incorporated into the primer 3’ terminus paired with template at the postinsertion site. Pre-steady-state kinetics was carried out using proteins without the D598A mutation. The D598A mutant crystallizes in a lattice that differs from the wild-type (but has the same space group) (37). In the wild-type lattice, the O helix is constrained in the open conformation by a lattice contact; in the D598A, it is not (32), and therefore it can move freely to form open, closed, and intermediate conformational states, as described below.

**Crystallization of Nucleotide Substrates and DNA Complexes**

Oligonucleotides were synthesized by Midland Certified Reagent Co. (Midland, TX) at GF grade. Ternary complexes were crystallized using 5’-CTGACTC<sub>α</sub>C-C-3’ as primer (dideoxynucleotide terminator traps the complex before chemistry) and 5’-CATGGGATCAGG-G-3’ as template. The primer and template strands were annealed as described (32). Ultrapure rCTP and ddCTP were purchased from U.S. Biochemical Corp., and dCTP was purchased from Promega (Madison, WI). Protein, primer-template duplex (protein/DNA, 1:3 molar ratio), nucleotides (dCTP or ddCTP, 10 mM; rCTP, 30 mM; Mg<sup>2+</sup> or Mn<sup>2+</sup>, SO<sub>4</sub><sup>2-</sup> salts; nucleotide/metal, 1:2 ratio) were incubated (1 h, room temperature) and crystallized as described previously (32). rCTP was incorporated (protein/DNA, 1:3 molar ratio; 30 mM rCTP; 60 mM MnSO<sub>4</sub>; 1 h, room temperature) and crystallized in a binary complex using a 5’-GCGAT-CACGTA-3’ (primer) and 5’-GACGTACGTGATCGCA-3’ (template) DNA substrate as described (32).

**Nucleoside Sugar Specificity of a DNA Polymerase**

All crystals were flash frozen in liquid nitrogen directly out of the crystallization drop. Data were collected at SIBYLS and SER-CAT beamlines. All crystals belong to the space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>. The closed form crystals (complexes 1–8) contain two molecules in the asymmetric unit, whereas the closed form complex 9 and the open form complex 10 contain one. Complex 8 was processed using HKL2000 (42), and all others were processed using XDS (43). Complexes 1–6 exhibit ice ring formation; accordingly, the reflections in shells 2.28–2.22 Å and 1.948–1.888 Å were excluded.

**Structure Determination**

Ternary Complexes 1–8—A closed form ternary complex (PDB code 2HVI (37)) from which the active site residues on the O helix and preceding loop region (residues 681–727), incoming nucleotide substrate, metal ions, surface-bound sulfate (from the crystallization solution), and water molecules had been removed was used as the starting model for solving the structures of complexes 1, 2, 4, 5, and 7. Complexes 3, 6, and 8 are the Mg<sup>2+</sup> or Mn<sup>2+</sup> counterparts for complexes 4, 5, and 7. Preliminary refinement showed that the two complexes with different metal ions are almost identical. Therefore, fully built and refined structures of complexes 4, 5, and 7 with water molecules removed were used as the starting models for their metal-substituted counterparts.

Initial electron density maps for complexes 1, 2, 4, 5, and 7 were generated by Fourier synthesis after rigid body refinement using the modified starting model 2HVI in PHENIX (44). Missing residues, nucleotide substrates, metal ions, and small molecules for all complexes were manually built in Coot (45) into the electron density map generated by iterative rounds of model building and refinement with PHENIX (44). All residues in the complexes were evaluated by visual inspection to fit the map. Simulated annealing refinement was then applied to further improve the model. TLS groups were determined using PHENIX find_tls_groups (44), and TLS parameters were refined subsequently. Water molecules were added automatically in the refinement followed by visual inspection.

For the closed form ternary complexes 1–8, the two molecules in the asymmetric unit are not the same; molecule 1 (chains D, E, and F in the PDB) is more ordered than molecule 2 (chains A, B, and C) (chain naming follows previously published structures (37)). Each molecule in complexes 1, 2, 7, and 8 adopts a single unique conformation (closed, ajar, or open). In complexes 3 and 4, molecule 1 does not adopt a unique conformation (molecule 2 is open and empty), and it was modeled initially as a mixture of two distinct ajar O helix conformations. After refinement, additional densities around the O helix could be observed in difference maps, suggesting that there are additional conformations that have not been accounted for in the model. At least two distinct rCTP conformations were observed. In the final, deposited models for complexes 3 and 4, the conformationally diverse portions of the O helix, its preceding loop region, and rCTP were omitted. For complexes 5 and 6,
molecule 1 adopted a single ajar conformation with bound rCTP, whereas molecule 2 adopted an ensemble of conformations similar to complexes 3 and 4; in the deposited structures for complexes 5 and 6, these also were omitted from the model. *Closed Form Binary Complex 9—*The E658A mutant, DNA, dCTP, and Mg$^{2+}$ complex was crystallized as described above for the closed form ternary complexes 1–8. However, the unit cell dimensions of this complex resemble those observed for an open, binary crystal form (31, 32), and it contains only one molecule in the asymmetric unit. The structure was determined by molecular replacement with Phaser (46), using as the search model the open, empty molecule of complex 7, which carries the same mutation and contains a DNA duplex with the same sequence. The output model was refined with simulated annealing in PHENIX (44) followed by iterative rounds of visual inspection, manual building, and refinement as described above. dCTP was not observed in the active site of this complex but was observed to bind at a remote surface site (37–39), confirming its presence in the crystal.

*Open Form Binary Complex 10—an open form binary complex (PDB code 1L5U) with the base pair positioned at the pol/strand insertion site surface-bound sulfate, and water molecules removed. The model for the rC nucleotide incorporated at the primer 3’ terminus was built into the electron density map. The structure was refined as described above. There is one molecule in the asymmetric unit with a single, open conformation.*

**Structure Quality Checks**

For all structure determinations, free reflections were generated in XDSCONV (43) or CCP4 (47) by combining inherited free reflections from the starting model and 5% randomly selected reflections beyond the resolution of the starting model for each complex. All structures were refined to good protein geometry with no Ramachandran outliers and ~98% residues in the Ramachandran favored region according to MolProbity (48). Data, refinement statistics, Ramachandran plot summary, and Protein Data Bank accession codes are presented in Table 3. Composite omit maps were generated in CNS (49). Figures and superpositions (all protein residues used) were prepared in PyMOL (Schrödinger, LLC). [Nucleotide Sugar Specificity of a DNA Polymerase](#)

### Table 2

| Protein | $K_a$ (μM) | $K_{mut}$ (μM) | Selectivity$^a$ |
|---------|------------|----------------|-----------------|
| WT      | 33.2       | 1,800          | 24,000          |
| E658A   | 16.2       | 753            | 170             |
| F710Y   | 36.6       | 674            | 46,000          |

$^a$ Defined as the ratio of $K_{pol}/K_a$ for dCTP versus rCTP incorporation opposite template dG.

---

**Nucleotide Incorporation Selectivity—**Nucleotide affinities and incorporation rates of dCTP, rCTP, and ddCTP were determined for wild-type BF and its E658A and F710Y mutants by pre-steady-state kinetics (Table 2). These values are consistent with those determined for other DNA polymerases in the A family (14, 19, 50, 51). Wild-type BF favors dCTP incorporation over rCTP or ddCTP by 4 and 3 orders of magnitude, respectively. The E658A mutant decreases the selectivity against rCTP by 2 orders of magnitude, primarily as a consequence of a 100-fold loss in the dCTP catalytic rate, consistent with the behavior of this mutation in other members of the DNA polymerase A family (Fig. 1). The F710Y mutation renders incorporation of ddCTP essentially as efficient as dCTP and slightly boosts the catalytic rate of the latter. This mutation also improves binding of rCTP but compromises its incorporation rate to a similar degree.

**Nucleotide Positioning, O Helix Motion, and Water Structure—**High resolution structures of dCTP, rCTP, and ddCTP complexes were obtained for the wild-type active site BF and the F710Y and E658A mutants using a primer lacking a 3’-OH of its terminus (Table 3). In all crystallized ternary complexes, a surface D598A mutation was included to destabilize a crystal contact that favors the open state in the wild-type crystal lattice (32). The resulting crystal form contains two molecules in the asymmetric unit. In all complexes, molecule 1 is more ordered than 2; accordingly, our comparative analysis is based on molecule 1. In all complexes, the nucleotide is positioned at the insertion site and forms hydrogen bonds with dG on the template strand. Mg$^{2+}$ and Mn$^{2+}$ were used in the rCTP complexes.
to determine the location of bound metal by anomalous scattering of the latter.

In the wild-type active site, the rCTP (Fig. 2) and ddCTP (Fig. 3) nucleotide triphosphates are displaced from the dCTP position. Consequently, the rCTP or ddCTP α-phosphates are not aligned properly with the primer strand terminus, consistent with the diminished $k_{\text{cat}}$ rate of these nucleotides relative to dCTP. The structure of the rCTP complex with wild-type BF is complicated by the presence of a mixture of two rCTP conformations, both with misaligned α-phosphates; this situation is simplified in the rCTP complex of the F710Y variant (which retains a strong preference for dCTP over rCTP; Table 2) in which only one, misaligned, misconfiguration is observed (Fig. 2).

In all misaligned rCTP or ddCTP complexes, the O helix fails to adopt the fully closed conformation (Figs. 2 (A and E) and 3A). Furthermore, these complexes contain a water layer that is absent in the fully closed wild-type dCTP and F710Y ddCTP complexes (Fig. 4). A water layer also is present in the fully open (e.g. PDB code 1L3S (32)) and ajar conformations (PDB codes 3HP6 (38) and 3PX4 (39)); its displacement is therefore part of the mechanism by which improper nucleotides are selected against.

$\text{ddCTP}$—The failure of ddCTP to displace the water layer present in the open state, with the accompanying lack of O helix closure and misalignment of the α-phosphate, is accompanied by the loss of two hydrogen bonds, formed between dCTP 3′-OH and its own β-phosphate oxygen and the backbone amide of Glu-658 (Fig. 3, C and D). These lost hydrogen bonds are compensated for in the F710Y mutant, which is located on the O helix surface facing the bound nucleotide. In the F710Y-ddCTP complex, the mutant phenolic hydroxyl substitutes for the 3′-OH to interact with the β-phosphate of bound ddCTP and backbone amide of Glu-658 (Fig. 3G). As a consequence, the ddCTP affinity improves sufficiently (~40-fold; Table 2) to displace the water layer and close the O helix (Figs. 3 (E and F) and 4C). The F710Y-ddCTP complex is essentially indistinguishable from the wild-type dCTP complex, accounting for their nearly identical incorporation rates. We note that Tyr-710 is a natural variant in a number of family A DNA polymerases (e.g. T7 DNA polymerase) (52).

$r\text{CTP}$—In the wild-type enzyme in the presence of Mg$^{2+}$ or Mn$^{2+}$, rCTP adopts an ensemble of conformations, and the O helix exhibits an ensemble of states, none of which correspond to the fully closed form. This ensemble is simplified in the F710Y mutant in which both the bound rCTP and the O helix adopt a single, misaligned, non-closed conformation with either metal (Fig. 2A). In this complex, four effects are observed that misalign the α-phosphate and interfere with nucleotide incorporation. First, rather than forming a steric block, the Glu-658 carboxylate interacts with the rCTP 2′-OH (2.7 Å apart) (Fig. 2B). Consequently, the position of the entire rCTP is displaced from that of dCTP to such an extent that even the template dG base moves to retain pairing between the two bases, albeit with base pair parameters that deviate substantially from those of canonical Watson-Crick base pairs (Table 4). Furthermore, the 2′-OH forms a hydrogen bond with Arg-615, which normally forms a hydrogen bond with the sugar O4 atom (Fig. 2B). This residue also participates in recognizing cognate base pairs by reading out their minor groove hydrogen bonding pattern (31). Second, the rCTP displacement opens an aperture between the bound nucleotide and the wall of the active site opposite the O helix, which is filled with a water layer that adopts a structure different from what was observed in the ddCTP complex (Fig. 4D). Third, the triphosphate binds in an altered conformation that is stabilized by this water layer (Fig. 5, A and B). The absence of an anomalous scattering signal revealed that in this conformation, the Mg$^{2+}$ (substituted with Mn$^{2+}$ to obtain an anomalous signal) that normally assembles with the nucleotide is absent and replaced by two water molecules (Fig. 2C). The direct hydrogen bond between the 3′-OH and its own β-phosphate oxygen observed in the dCTP complex is replaced by the Tyr-710 phenolic hydroxyl (Fig. 2C). This cyclic, non-covalent interaction within the nucleotide triphosphate has also been observed in ternary complexes of B-, C-, X-, and Y-family DNA polymerases and reverse transcriptase family (Fig. 6). Finally, the direct interaction between 3′-OH and Glu-658 main-chain amide is replaced by a water-mediated hydrogen bond (Fig. 2C). The stabilizing interaction of the bound rCTP β-phosphate with Tyr-710 on the O helix and the Glu-658 backbone amide via a water molecule presumably together improve affinity sufficiently (Table 2) to select a single conformation of the ensemble present in the wild-type complex.

The loss of the interaction between the glutamate carboxylate and 2′-OH of the rCTP complex in the E658A single mutant does not suffice to form the closed state (Fig. 2E). The diminished selectivity between rCTP and dCTP (Table 2) is primarily a consequence of a 100-fold loss of activity for dCTP.

$d\text{CTP}$—To elucidate the structural mechanism for the loss of dCTP incorporation in the E658A mutant, we co-crystallized this mutant with dCTP and Mg$^{2+}$. The resulting complex (Fig. 7) contains only one molecule in the asymmetric unit, compared with the two observed in all other ternary complexes. This complex does not bind a dCTP at its active site (dCTP is observed to bind at a remote surface site, confirming its presence in the crystal). The polymerase conformation is similar to...
TABLE 3
Crystallographic data collection and refinement statistics

In complexes 1–8, the nucleotide substrate is trapped at the insertion site (n) prior to incorporation, using a primer strand with a dideoxynucleotide terminus. These proteins contain the D598A mutation that destabilizes a crystal contact (32). Active site mutations are indicated. Complexes 3 and 4 exhibit an ensemble of rCTP and O helix conformations; their components were omitted from the models. Analysis of rCTP binding is therefore based on complexes 5 and 6. In complex 9, the protein also carries the D598A mutation. The active site is empty, but there is a dCTP bound to the surface. In Complex 10, the ribonucleotide has been incorporated at the primer terminus and is positioned in the postinsertion site (n/H11002).

| Data collection | 1, dCTP-dG | 2, ddCTP-dG | 3, rCTP-dG | 4, rCTP-dG | 5, F710Y-rCTP-dG | 6, F710Y-rCTP-dG | 7, E658A-rCTP-dG | 8, E658A-rCTP-dG | 9, E658A-DNA | 10, WT-rC |
|-----------------|-----------|------------|-----------|-----------|-----------------|-----------------|-----------------|----------------|-------------|---------|
| Space group     | P 2 1 2 1 | P 2 1 2 1  | P 2 1 2 1 | P 2 1 2 1  | P 2 1 2 1       | P 2 1 2 1       | P 2 1 2 1       | P 2 1 2 1       | P 2 1 2 1    | P 2 1 2 1 |
| Cell dimensions (Å) |          |            |           |           |                 |                 |                 |                 |             |         |
| a               | 93.7      | 94.2       | 93.9      | 93.8      | 93.9            | 93.7            | 93.5            | 93.0            | 87.9         | 88.4    |
| b               | 109.4     | 109.3      | 108.9     | 108.8     | 109.5           | 109.1           | 108.0           | 108.8           | 93.8         | 93.4    |
| c               | 150.3     | 149.9      | 151.0     | 150.7     | 150.6           | 150.1           | 149.9           | 149.8           | 104.9        | 105.8   |
| No. of complexes in asymmetric unit | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 |         |
| Wavelength      | 1.0000    | 1.0000     | 1.1599    | 1.1580    | 1.1599          | 1.1519          | 1.1519          | 1.1519          | 1.1519       | 1.1519   |
| Resolution (Å)  | 100-1.69  | 100-1.74   | 100-1.68  | 100-1.68  | 100-1.67        | 100-1.66        | 100-1.66        | 100-1.66        | 100-1.66     | 100-1.66 |
| Outer shell (Å) | 1.79-1.69 | 1.84-1.74  | 1.78-1.68 | 1.78-1.68 | 1.77-1.67       | 1.76-1.66       | 1.69-1.59       | 1.98-1.95       | 1.67-1.58    | 1.78-1.66 |
| Rsym (%)        | 6.6 (49.1) | 7.7 (49.0) | 6.2 (46.6) | 6.7 (48.0) | 5.8 (49.3)      | 5.9 (49.1)      | 5.7 (47.8)      | 9.2 (68.7)      | 5.9 (46.7)   | 4.1 (48.9) |
| Completeness (%)| 90.2 (94.9)| 89.8 (100)| 89.6 (97.2)| 90.9 (94.4)| 91.2 (99.8)     | 90.6 (93.4)     | 98.7 (92.3)     | 94.5 (100)      | 98.8 (98.9)  | 99.8 (99.9) |
| Redundancy      | 5.2 (50.0)| 7.2 (74.0)| 9.9 (5.9) | 10.8 (6.2) | 9.1 (5.6)       | 9.1 (5.4)       | 9.4 (5.9)       | 6.2 (6.5)       | 10.0 (5.5)   | 4.8 (4.8) |

Refinement

| Resolution (Å)  | 28.8-1.69 | 29.4-1.74 | 34.4-1.68 | 64.3-1.68 | 79.7-1.67       | 71.1-1.66        | 79.4-1.60        | 156-1.16        | 156-1.16     | 54.7-1.58 |
| No. of reflections | 148,420 | 138,100   | 152,367   | 153,529   | 156,214         | 156,479          | 192,914          | 121,902         | 103,589     | 45.4-1.66 |
| Rwork (%)       | 15.8       | 16.7       | 17.6       | 17.4       | 17.2            | 17.1             | 17.4             | 17.7             | 16.0         | 16.0-18.1 |
| Root mean square deviations (Å) | 0.011 | 0.012 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.014 | 0.010 | 0.010 |
| Bond lengths (Å) | 0.011 | 0.012 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.014 | 0.010 | 0.010 |

| Root mean square deviations (Å) | 0.011 | 0.012 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.014 | 0.010 | 0.010 |
| Bond angles (degrees) | 1.347 | 1.395 | 1.345 | 1.342 | 1.374 | 1.379 | 1.334 | 1.335 | 1.291 | 1.395 |
| Ramachandran plot (%) | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| PDB code | 4DQI | 4DQP | 4DS5 | 4DS4 | 4DSE | 4DSF | 4DQQ | 4DQR | 4DOD | 4DQS |

Numbers in parentheses correspond to values in the outer resolution shell.

Nucleoside Sugar Specificity of a DNA Polymerase

JOURNAL OF BIOLOGICAL CHEMISTRY
VOLUME 287 • NUMBER 34 • AUGUST 17, 2012
the open state (31, 32), but the C-terminal segment of the O helix (residues 707–714) has bent toward the primer DNA 3' terminus (Fig. 7A). A conserved water molecule that usually is present both in open structures (31, 32) (Fig. 7C) and in ternary complexes (37–39) is displaced by the backbone carbonyl of Val-713 (Fig. 7B). It also is present in all structures except this one, reported here. This water molecule normally interacts with the minor groove of the unincorporated dNTP-template.
base pair. Its hydrogen-bonding interactions are likely to be sufficiently strong that they perturb the intrinsic tautomeric equilibria of the free nucleotides (39). Loss of this water molecule can therefore be expected to have a significant impact on dNTP incorporation rates.

The loss of this important water molecule in the E658A mutant is accompanied by several rearrangements in the active site, which are compensated for by the 3′-OH in the rCTP complex. Normally, Gln-797, Asn-793, and Glu-658 form a hydrogen-bonded network that coordinates this water. This network is broken in E658A such that Gln-797 adopts a different side chain rotamer conformation, and the Val-713 carbonyl is displaced (Fig. 7D). These motions together destroy the water coordination sphere. However, in the rCTP complex, the 3′-OH is able to compensate for the loss of the carboxylate in the E658A, because in the properly formed closed state, this hydroxyl occupies the position vacated by the carboxylate, restoring the hydrogen-binding network (Fig. 4E). Despite the restoration of this critical water in the rCTP complex, the incorporation rate of this nucleotide remains low because it still adopts distorted sugar and triphosphate conformations (Figs. 2F and 5, C and D). The direct interaction between the 3′-OH and β-phosphate is replaced by water; there is no interaction between the 3′-OH and the E658A backbone amide; and the triphosphate binds Mg^{2+} (confirmed by Mn^{2+} anomalous signal) but not at full occupancy (Fig. 2G). The Ω helix remains open to a similar degree as observed for the F710Y-rCTP complex.

Base Pairing of the Incorporated rCTP—Covalently incorporated, cognate base pairs positioned at the postinsertion site form hydrogen bonds between their minor groove N3 (purines) or O2 atoms (pyrimidines) and Arg-615 and Gln-797 (31). We incorporated rCTP at the primer strand 3′ terminus by catalysis in solution in the presence of Mn^{2+} (53) prior to crystallization. This binary complex between wild-type BF and DNA was crys-

### Table 4

Base pair parameters of ribonucleotide captured at the insertion site

| Base pair            | \( \lambda_{\text{primary}} \) | \( \lambda_{\text{template}} \) | \( d_{\text{C1'-C1}} \) | Shear | Stretch | Stagger | Buckle | Propeller | Opening |
|----------------------|-------------------------------|-------------------------------|-------------------------|-------|---------|---------|--------|-----------|---------|
| F710Y-rCTP:dG (Mg^{2+}) | 54.0                          | 52.3                          | 10.7                    | 0.32  | -0.16   | 0.24    | 5.54   | -19.00    | 0.63    |
| E658A-rCTP:dG (Mg^{2+})  | 59.7                          | 54.7                          | 10.6                    | 0.37  | -0.09   | -0.30   | 8.27   | -4.13     | 0.77    |
| Watson-Crick          | 57.8 ± 1.8                    | 54.7 ± 0.9                    | 10.6 ± 0.1              | 0.06  | ±0.28   | -0.10 ± 0.05 | -0.09 ± 0.08 | 4.31 ± 4.5 | -6.88 ± 3.41 | 1.86 ± 0.80 |

\( \lambda_{\text{primary}} \) and \( \lambda_{\text{template}} \) are defined as the angle between the glycosidic bond of primer or template nucleotide and the line drawn between the C1′ atoms of the base pair. \( d_{\text{C1'-C1}} \) is the distance between the C1′ atoms of the base pair. All other base pair parameters are defined (66). All values were calculated in 3DNA (67).

### Figure 5

**Nucleotide triphosphate conformations.** A–C, individual; D, superimposed. Gray, wild-type active site-dCTP; blue, F710Y-rCTP; pink, E658A-rCTP. Two views are shown.
tallized in the absence of additional nucleotide. The resulting crystals contain one molecule in the asymmetric unit in the open form (Table 3). The rC is positioned in the postinsertion site and retains the minor groove readout by Arg-615 and Gln-797 observed for deoxynucleotide (31) (Fig. 8), emphasizing the importance of selecting against rNTPs prior to their incorporation.

**DISCUSSION**

The steric gate hypothesis postulates that discrimination between dNTPs and rNTPs is primarily the consequence of a steric blockage imposed by a single DNA polymerase residue on the rNTP 2'-OH (28). Although reducing the bulk of this residue by mutation improves the incorporation rate of rNTPs relative to dNTPs, it invariably decreases dNTP incorporation (Fig. 1). Furthermore, dNTP remains favored over rNTP incorporation in these mutants. These observations indicate that the “steric gate” residue functions not as a simple block but also affects the mechanism of dNTP incorporation. The structural observations reported in this study suggest that the origins for this complex behavior lie within the dynamics of O helix motions (32, 38, 39).

We were able to capture rNTPs and ddNTPs at the “insertion site,” paired with the template DNA, prior to reacting with the primer terminus and rNTP following its incorporation at the
“postinsertion site.” In the sugar moiety discrimination process, the O helix motions are interfered with, resulting in non-productive conformations that differ from those observed for mismatches or lesions (37–39). Water-mediated interactions, which are absent in the closed state, contribute to the formation of the binding sites that trap the non-cognate nucleotides in the intermediate O helix conformations. These sites are placed in similar locations as observed for cognate nucleotide complexes but differ in their interactions between the polymerase and bound nucleotide.

The O helix motions are therefore not simply a motion that distinguishes open and closed states, in which only the closed state binds NTPs. Instead, the motions set up an ensemble of subtle “traps” that ensnare non-cognate NTPs. The binding of the water molecules that contribute to shaping the ensemble of non-cognate binding sites is affected by polymerase mutants, thereby altering the thermodynamic landscape of the non-cognate nucleotide traps. This mechanism accounts for the large variation in specificities that has been noted (7, 54) and is consistent with a model based on the analysis of enzyme kinetic data in which specificity is a consequence of subtly different reaction pathways corresponding to the states particular to individual (mis)matched base pairs (2).

The ddCTP and rCTP complexes presented here illustrate these effects. The two water layers observed in the rCTP and ddCTP complexes (Fig. 4) differ from each other, stabilizing aberrant interactions (or lack of interaction) specific to the respective nucleotide. In the case of the rCTP complex in the F710Y enzyme, the interactions with water are sufficiently strong that they prevent binding of Mg\(^{2+}\) to the triphosphate. The remodeling of water structure in response to mutations is illustrated in the F710Y complex with ddCTP, in which the mutant Tyr-710 phenolic hydroxyl replaces a bound water such that the closed state binding becomes dominant, eliminating the selectivity against ddNTP incorporation.

Water exclusion in the active site has long been proposed to interfere with O helix motion and proper closed state assembly are distinct from base pair recognition, some polymerase residues play a role in both processes. This duality is responsible for the loss of dNTP incorporation in the E658A steric gate mutant and also involves water-mediated effects. Glu-658 forms part of a hydrogen-bonding network that holds in place a water molecule critical for reading out minor groove interactions between the NTP and the DNA template. In the E658A mutation, these interactions are destroyed, leading to the loss of the water and reduced dNTP incorporation. This loss is partially compensated for in the rNTP complex of this mutant, in which the 2’-OH takes the place of the missing carboxylate.

The subtle energetics generated by dynamically interconverting ensembles of conformational states is key to understanding protein folding (56), enzyme kinetics (2, 57, 58), and allostery (59, 60). The mechanistic principles established by this recently emerged view of protein structure-function relationships are central to understanding DNA polymerase fidelity. Here we show structural features that lead to the creation of such thermodynamic landscapes via the interplay between protein dynamics, ligand binding, and solvent interactions.

Acknowledgments—We thank S. Armstrong for assistance with site-directed mutagenesis and protein preparation, S. Langdon for assistance with capillary electrophoresis for fluorescence detection, and P. Modrich for use of the rapid quench flow instrument. Data for this study were collected at SIBYLS and SER-CAT (22-BM) beamlines.

REFERENCES

1. Rothwell, P. J., and Waksman, G. (2005) Structure and mechanism of DNA polymerases. Adv. Protein Chem. 71, 401–440
2. Joyce, C. M., and Benkovic, S. J. (2004) DNA polymerase fidelity. Kinetics, structure, and checkpoints. Biochemistry 43, 14317–14324
3. Steitz, T. A., and Yin, Y. W. (2004) Accuracy, lesion bypass, strand displacement and translocation by DNA polymerases. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359, 17–23
4. Johnson, K. A. (1993) Conformational coupling in DNA polymerase fidelity. Annu. Rev. Biochem. 62, 685–713
5. Kool, E. T. (2002) Active site tightness and substrate fit in DNA replication. Annu. Rev. Biochem. 71, 191–219
6. Echols, H., and Goodman, M. F. (1991) Fidelity mechanisms in DNA replication. Annu. Rev. Biochem. 60, 477–511
7. Kunkel, T. A., and Bebenek, K. (2000) DNA replication fidelity. Annu. Rev. Biochem. 69, 497–529
8. Traut, T. W. (1994) Physiological concentrations of purines and pyrimidines. Mol. Cell Biochem. 140, 1–22
9. Ferraro, P., Franzolin, E., Pontarini, G., Reichard, P., and Bianchi, V. (2010) Quantitation of cellular deoxyxynucleoside triphosphates. Nucleic Acids Res. 38, e85
10. Nick McElhinny, S. A., Watts, B. E., Kumar, D., Watt, D. L., Lundström, E. B., Burgers, P. M., Johansson, E., Chabes, A., and Kunkel, T. A. (2010) Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. Proc. Natl. Acad. Sci. U.S.A. 107, 4949–4954
11. Brown, J. A., and Suo, Z. (2011) Unlocking the sugar “steric gate” of DNA polymerase I. J. Mol. Biol. 413, 1135–1142
12. Cricc, F. (1970) Central dogma of molecular biology. Nature 227, 561–563
13. Nick McElhinny, S. A., Kumar, D., Clark, A. B., Watt, D. L., Watts, B. E., Lundström, E. B., Johansson, E., Chabes, A., and Kunkel, T. A. (2010) Genome instability due to ribonucleotide incorporation into DNA. Nat. Chem. Biol. 6, 774–781
14. Asatke, M., Ng, K., Grindley, N. D., and Joyce, C. M. (1998) A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment)
from incorporating ribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3402–3407

15. Lewis, W., Day, B. J., and Copeland, W. C. (2003) Mitochondrial toxicity of NRTI antiviral drugs. An integrated cellular perspective. *Nat. Rev. Drug Discov.* **2**, 812–822

16. Henry, A. A., and Romesberg, F. E. (2005) The evolution of DNA polymerases with novel activities. *Curr. Opin. Biotechnol.* **16**, 370–377

17. Holmberg, R. C., Henry, A. A., and Romesberg, F. E. (2005) Directed evolution of novel polymerases. *Biometal Rev.* **22**, 39–49

18. Patel, P. H., and Loeb, L. A. (2000) Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J. Biol. Chem.* **275**, 40266–40272

19. Kirouac, K. N., Suo, Z., and Ling, H. (2011) Structural mechanism of ribonucleotide discrimination and reverse transcription by the human mitochondrial DNA polymerase. *J. Biol. Chem.* **286**, 31490–31500

20. Gao, G., Orlova, M., Georgiadis, M. M., Hendrickson, W. A., and Goff, S. P. (1997) Conferring RNA polymerase activity to a DNA polymerase. *J. Biol. Chem.* **272**, 10179–10186

21. NRTI antiviral drugs. An integrated cellular perspective. *Methods Enzymol.* **276**, 307–326

22. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Wilson, S. H. (2011) Molecular insights into DNA polymerase determinants for ribonucleotide insertion. *J. Biol. Chem.* **286**, 31650–31660

23. Cavanaugh, N. A., Beard, W. A., Batra, V. K., Perera, L., Pedersen, L. G., and Wilson, S. H. (2011) Crystal structure of a thermostable Bacillus subtilis DNA polymerase crystal. *Nature* **475**, 2126–2132

24. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system. A new software suite for macromolecular structure determination. *Acta Cryst.* **D54**, 905–921

25. Astatke, M., Grindley, N. D., and Joyce, C. M. (2003) An error-prone family Y DNA polymerase (DinB homolog from *Sulfolobus solfataricus*) uses a “steric gate” residue for discrimination against ribonucleotides. *Nucleic Acids Res.* **31**, 4129–4137

26. Joyce, C. M. (1997) Choosing the right sugar. How polymerases select a nucleotide substrate. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1619–1622

27. Sawyer, L. A., and Goodman, M. F. (1986) Comparison of nucleotide interactions in water, proteins, and vacuum. Model for DNA polymerase fidelity. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6339–6343

28. Johnson, S. J., Taylor, J. S., and Beece, L. S. (2003) Processive DNA synthesis observed in a Y-family DNA polymerase. *J. Mol. Biol.* **236**, 577–586

29. Minnick, D. T., Bebenek, K., Osheroff, W. P., Turner, R. M., Jr., Astatke, M., Grindley, N. D., and Joyce, C. M. (2008) Energy landscape of enzyme catalysis. *J. Biol. Chem.* **283**, 6140–6148

30. McCulloch, S. D., and Kunkel, T. A. (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res.* **18**, 148–161

31. Minknick, D. T., Bebenek, K., Osheroff, W. P., Turner, R. M., Jr., Astatke, M., Liu, L., Kunkel, T. A., and Joyce, C. M. (1999) Side chains that influence fidelity at the polymerase active site of *Escherichia coli* DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **274**, 3067–3075

32. Van de Sande, J. H., Loewen, P. C., and Khorana, H. G. (1972) Studies on ribonucleotides. 118. A further study of ribonucleotide incorporation into deoxyribonucleic acid chains by deoxyribonucleic acid polymerase I of *Escherichia coli*. *J. Biol. Chem.* **247**, 7514–7525

33. Petruska, J., Sowers, L. C., and Goodman, M. F. (1986) Comparison of nucleotide interactions in water, proteins, and vacuum. Model for DNA polymerase fidelity. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1559–1562

34. McLock, S. D., and Kunkel, T. A. (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res.* **18**, 148–161

35. Smock, R. G., and Giersch, L. M. (2009) Sending signals dynamically. *Science* **324**, 198–203

36. Benkovic, S. I., Hammes, G. G., and Hammes-Schiffer, S. (2008) Free energy landscape of enzyme catalysis. *Biochemistry* **47**, 3317–3321

37. Nashine, V. C., Hammes-Schiffer, S., and Benkovic, S. I. (2010) Coupled
Nucleotides and enzyme catalysis. *Curr. Opin. Chem. Biol.* 14, 644–651
59. Goodey, N. M., and Benkovic, S. J. (2008) Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* 4, 474–482
60. Reynolds, K. A., McLaughlin, R. N., and Ranganathan, R. (2011) Hot spots for allosteric regulation on protein surfaces. *Cell* 147, 1564–1575
61. Wang, M., Xia, S., Blaha, G., Steitz, T. A., Konigsberg, W. H., and Wang, J. (2011) Insights into base selectivity from the 1.8 Å resolution structure of an Rb69 DNA polymerase ternary complex. *Biochemistry* 50, 581–590
62. Evans, R. J., Davies, D. R., Bullard, J. M., Christensen, J., Green, L. S., Guiles, J. W., Pata, J. D., Ribble, W. K., Janjic, N., and Jarvis, T. C. (2008) Structure of PolC reveals unique DNA binding and fidelity determinants. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20695–20700
63. Batra, V. K., Beard, W. A., Shock, D. D., Krahn, J. M., Pedersen, L. C., and Wilson, S. H. (2006) Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure* 14, 757–766
64. Bierstümpfel, C., Zhao, Y., Kondo, Y., Ramón-Maíques, S., Gregory, M., Lee, J. Y., Masutani, C., Lehmann, A. R., Hanaoka, F., and Yang, W. (2010) Structure and mechanism of human DNA polymerase η. *Nature* 465, 1044–1048
65. Lansdon, E. B., Samuel, D., Lappacan, L., Brendza, K. M., White, K. L., Hung, M., Liu, X., Boojamma, C. G., Mackman, R. L., Cihlar, T., Ray, A. S., McGrath, M. E., and Swaminathan, S. (2010) Visualizing the molecular interactions of a nucleotide analog, GS-9148, with HIV-1 reverse transcriptase-DNA complex. *J. Mol. Biol.* 397, 967–978
66. Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C. S., von Kitzing, E., Wang, A. H., and Zhurkin, V. B. (1989) Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* 205, 787–791
67. Lu, X. J., and Olson, W. K. (2003) 3DNA. A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 31, 5108–5121