Synthesis of phosphoramidate-linked DNA by a modified DNA polymerase

Victor S. Lelyveld[a,b], Wen Zhang[a,b,1], and Jack W. Szostak[a,b,c,d,2]

[a]Department of Molecular Biology, Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA 02114; [b]Department of Genetics, Harvard Medical School, Boston, MA 02115; [c]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138; and [d]Howard Hughes Medical Institute, Massachusetts General Hospital, Boston, MA 02114

Edited by John Chaput, University of California, Irvine, CA, and accepted by Editorial Board Member Stephen J. Benkovic February 20, 2020 (received for review December 19, 2019)

All known polymerases copy genetic material by catalyzing phosphodiester bond formation. This highly conserved activity proceeds by a common mechanism, such that incorporated nucleoside analogs terminate chain elongation if the resulting primer strand lacks a terminal hydroxyl group. Even conservatively substituted 3′-amino nucleotides generally act as chain terminators, and no enzymatic pathway for their polymerization has yet been found. Although 3′-amino nucleotides can be chemically coupled to yield stable oligonucleotides containing N3′→P5′ phosphoramidate (NP) bonds, no such internucleotide linkages are known to occur in nature. Here, we report that 3′-amino terminated primers are, in fact, slowly extended by the DNA polymerase from *B. stearothermophilus* in a template-directed manner. When its cofactor is Ca^{2+} rather than Mg^{2+}, the reaction is fivefold faster, permitting multiple turnover NP bond formation to yield NP-DNA strands from the corresponding 3′-amino-2′,3′-dideoxynucleoside 5′-triphosphates. A single active site mutation further enhances the rate of NP-DNA synthesis by an additional 21-fold. We show that DNA-dependent NP-DNA polymerase activity depends on conserved active site residues and propose a likely mechanism for this activity based on a series of crystal structures of bound complexes. Our results significantly broaden the catalytic scope of polymerase activity and suggest the feasibility of a genetic transition between native nucleic acids and NP-DNA.

polymerases | alternative genetic polymers | NP-DNA | XNA

Eschenmoser proposed that, since cellular life is bounded by the rules of chemistry, judicious study of chemical reactions might yield models for how life could possibly emerge (1). If the chemical nature of the genetic material is conserved across all known life on Earth, it follows that the problem of abiogenesis might then be distilled down to a retrosynthetic question: What are the chemical origins of this material?

Orgel cautioned, however, that a view of Earth's cellular life arising as an "inevitable" consequence of chemistry and physics "should be resisted," and proper deference must be paid to the role of "historical accident" (2). It follows that scrutiny of alternative informational chemistries may shed light on the likelihood of such accidents occurring in chemical space. In pursuit of this question, Orgel (3) and Eschenmoser (4) began to explore the chemical neighborhood of RNA in search of genetic materials that could be competitors of—or alternatives to—those found in terrestrial cells.

Of particular interest, a conservative DNA analog containing N3′→P5′ phosphoramidate (NP) bonds has been intensively studied in a nonenzymatic pathway for copying information in model protocells (5). In the nonenzymatic system, 3′-amino-2′,3′-dideoxyribonucleotides activated as 5′-phosphorimidazolides can achieve high rates and extents of template-directed polymerization to form NP-linked DNA (NP-DNA) (6, 7) even when encapsulated (5).

Internucleotidyl N3′→P5′ linkages are not known to exist in nature, and no enzymatic synthesis of NP-DNA has yet been observed, to our knowledge. Letsinger observed that DNA containing the highly labile P3′→N5′ linkage regioisomer could be produced by *E. coli* DNA polymerase I from nucleotide analogs that already contained a P_α→N5′ phosphoramidate (8), but no N-P bond formation occurs in this reaction. Synthetic 3′-amino-2′,3′-dideoxyribonucleoside 5′-triphosphates (nNTPs) have long been considered inhibitors of primer extension catalyzed by polymerases (9, 10). In a pharmacological context, 3′-amino-nucleosides are the in vivo reduction products of the corresponding 3′-azido-nucleotides (e.g., 3′-azido-thymidine or AZT) widely deployed for antiretroviral therapy (11). Phosphoramidate bonds are nevertheless found in the natural world, e.g., in ligase adenylyl-lysine intermediates (12).

The absence of an NP-DNA polymerase presents a severe obstacle to studying the informational and functional aspects of this polymer. Its absence in known biology also begs the question of whether there might be fundamental aspects of chemistry that preclude its involvement in any abiogenesis. Here, we report that N3′→P5′ bond formation can indeed be catalyzed by a thermophilic DNA polymerase, indicating that template-directed NP-DNA synthesis is chemically feasible by both nonenzymatic and enzymatic chemistries.

Significance

Life on Earth depends on polymerases. These enzymes copy genetic information to produce the DNA and RNA strands at the core of the central dogma. Polymerases act by forming phosphodiester linkages to produce polynucleotide strands. While synthetic chemistry can generate a broad range of alternative genetic materials with unnatural linkages, polymerases have so far been limited to forming O-P bonds. Here, we show that, in fact, unnatural N-P bonds can also be formed by a modified DNA polymerase. This template-directed activity generates complementary strands linked by phosphoramidate (NP) esters, an alternative backbone linkage only known to exist in the laboratory. The emergence of NP-DNA polymerase activity implies the biochemical plausibility of alternative central dogmas for cellular life.

Author contributions: V.S.L. and J.W.S. designed research; V.S.L. and W.Z. performed research; V.S.L. and W.Z. analyzed data; and V.S.L. and J.W.S. wrote the paper.

Competing interest statement: The authors have filed for a patent related to this work.

This article is a PNAS Direct Submission. J.C. is a guest editor invited by the Editorial Board.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB), http://www.wwpdb.org/ (PDB ID codes 6UR2, 6UR4, 6UR9, and 6US5).

[1]Present address: Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202.

[2]To whom correspondence may be addressed. Email: szostak@molbio.mgh.harvard.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1922400117/-/DCSupplemental.

First published March 18, 2020.

Results

We recently reported that reverse transcriptases can recognize NP-DNA templates and synthesize a cDNA strand, suggesting that the structural homology between NP-DNA and native nucleic acids is sufficient to support polymerase activity (13). However, enzymatic synthesis of NP-DNA would require not only recognition of the genetic polymer, but also a novel chemistry at the active site. For such catalysis to occur, the 3′-amino terminal primer would, at a minimum, need to adopt a conformation related to that seen in the native reaction center. We therefore sought to investigate the structural consequences of 3′-amino sugar substitution at the terminus of a primer in an enzymatic context. We grew crystals of the large fragment of DNA polymerase I (BF) from the thermophilic bacterium *Bacillus stearothermophilus* (Bst, now classified as *Geobacillus*) complexed with a DNA duplex in which the primer contains a terminal 3′-amino-2′,3′-dideoxycytidine (nC) residue (Fig. 1). The crystal structure of the complex was solved by molecular replacement to a resolution of 2.25 Å (Fig. 1 and *SI Appendix*, Tables S1 and S2). The enzyme was found in a 1:1 stoichiometry with the bound duplex in the asymmetric unit, with the nC terminated primer aligned at the canonical phosphodiester bond-forming active site (Fig. 1*A*, 0 Complex). Although the crystal was grown in the presence of free 2′-deoxyguanosine 5′-triphosphate (dGTP) and Mg^{2+}, no ordered metal ion or nucleotide was apparent. The structure is consistent with an "open" conformation (14), in which the "O helix" of the fingers domain (698–714) is arrayed distally from the primer terminus, allowing significant solvent access to the reaction center (Fig. 1*B*). Density for the 3′-amino-2′,3′-dideoxyribose sugar of the terminal nC residue was consistent with a C3′-endo conformation, in concordance with the A-form geometry seen in crystallographic and NMR studies of NP-DNA duplexes (15–17) (Fig. 1*C*).

Under different conditions, a slow-growing crystal emerged after ∼45 d, yielding a 2.27-Å structure of an alternative view of the complex. While the solved structure was again consistent with an open conformation, it was found during refinement that the observed density for the bound duplex was best explained as a post-translocated product complex containing a covalently incorporated deoxyguanosine (dG) residue at the +1 position of the primer strand (+1 complex, Fig. 1 *D*–*F*). The density of the added dG is well-defined in the complex, as is the density corresponding to the +2 template deoxyadenosine (dA), now positioned in the proximal templating position in a pocket formed by the O and P helices of the fingers domain (Fig. 1*F*).

A primer extension reaction occurring on a timescale of weeks could reflect a background uncatalyzed rate or a process that occurs solely *in crystallo*. We therefore sought to establish whether NP bond formation occurs in solution and, if so, whether this reaction is in fact catalyzed by the polymerase. When incubated under conditions similar to the mother liquor in which extension was first observed crystallographically, slow extension of the 3′-amino primer could be observed in a polymerase-dependent manner on a timescale of days in solution (*SI Appendix*, Fig. S1). By high-resolution mass analysis of the reactions, we detected ions consistent with the NP-bonded +1 product with 3.4 ppm mass error and corresponding isotopic distribution, while no products of 3′-5′ exonucleolytic activity and/or extension could be detected (*SI Appendix*, Fig. S2).

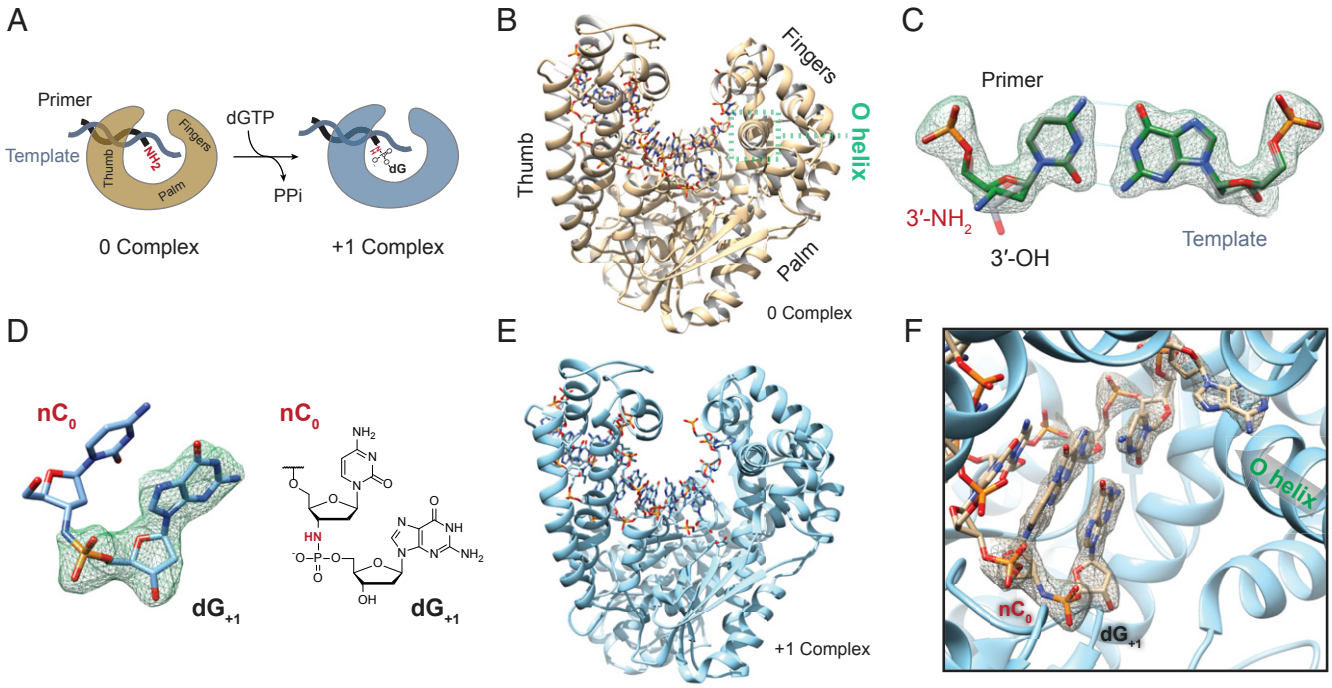

Fig. 1. Reactant and product polymerase "open" complexes containing a 3′-amino primer or a N3′→P5′ phosphoramidate bond. (*A*) Cartoon of ternary complexes. The "0 complex" contains WT BF, 3′-amino terminal DNA primer (5′-GCGATCAGnC, black), and DNA template (5′-ACACGCTGATCGCA, gray) in an open conformation. The "+1 complex" formed by in situ primer extension with dGTP, forming a phosphoramidate linkage, followed by translocation. (*B*) The 0 complex structure solved at 2.25 Å. BF (ribbon) is bound to unextended primer/template in the open conformation (PDB ID code 6UR4). (*C*) Overlay of the 3′-terminal nC sugar in the 0 complex (green bonds) and that of a dC terminated primer (gray, PDB ID code 1L5U) (26), as well as the 2Fo - Fc density map (green mesh, 2 σ) associated with the terminal nC:dG base pair. (*D*) Structure and omit map (green mesh, 4 σ) for the terminal dG (+1) residue extended in situ from the 3′-nC (0) terminated primer in the +1 complex and chemical structure as refined. (*E*) The +1 complex structure solved at 2.27 Å in an open posttranslocated conformation with +1 extended primer and template (PDB ID code 6UR2). (*F*) Enlarged view of the +1 complex active site, showing the extended dG residue and associated 2Fo - Fc map overlay (beige mesh, 2 σ).

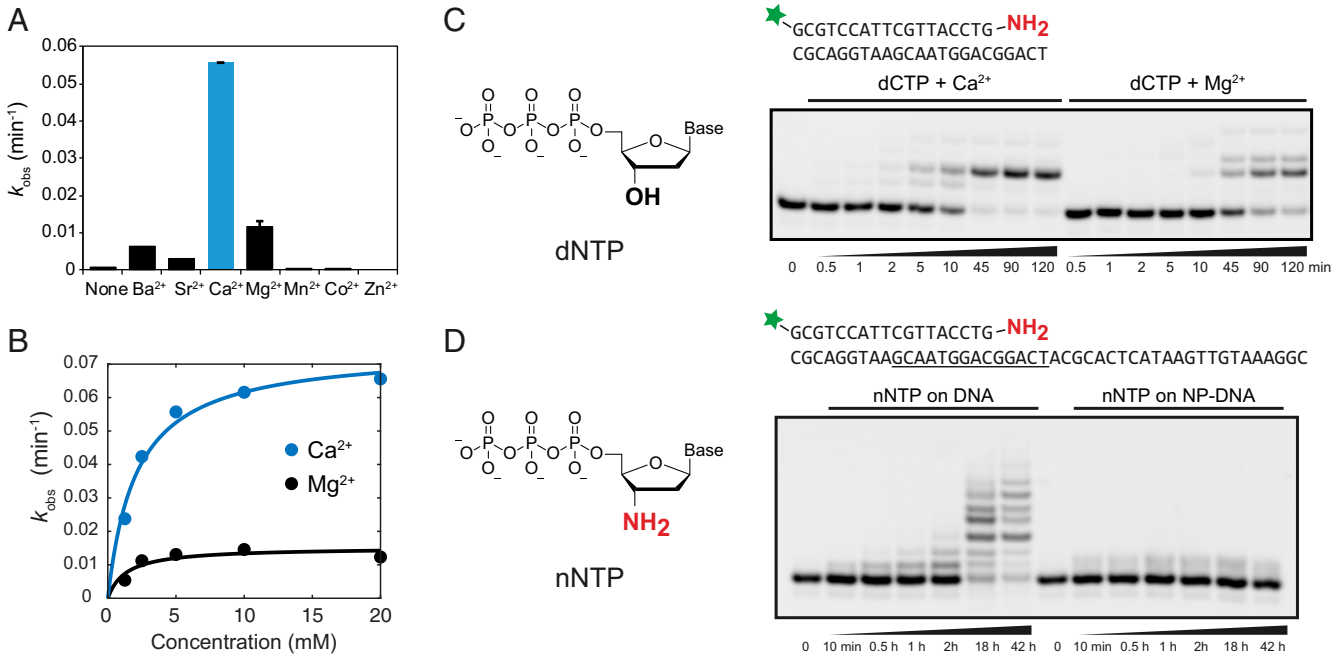

Fig. 2. Polymerase-catalyzed 3′-amino primer extension and NP-DNA synthesis. (*A*) Effect of divalent metal ion cofactors at 5 mM with 1 μM 3′-amino terminal DNA primer, 1.5 μM DNA template, and 1.1 μM WT BF at 55 °C in 40 mM Tris·HCl, pH 8.8, and 1 mM DTT (omitted for Co^{2+}). Error bars are SEM for $n = 2$. (*B*) Observed rate constant, k_{obs}, of 3′-amino primer extension by WT BF in the presence of 1 mM dCTP and varying concentrations of $CaCl_2$ (blue, $k_{pol} = 0.075$ min^{-1} and $K_{d,app} = 2.1$ mM) or $MgCl_2$ (black, $k_{pol} = 0.015$ min^{-1} and $K_{d,app} = 1.3$ mM). (*C*) Time course of 3′-amino primer extension reactions (*Right*) with 1 mM dCTP and 10 mM $CaCl_2$ or $MgCl_2$ visualized by denaturing PAGE. (*Upper*) Fluorescein-labeled 3′-amino terminal primer and template DNA duplex for *A*–*C*. (*D*) Chemical structure of nNTP substrates (*Left*), and time course of 3′-amino primer extension reactions with 1 mM nNTPs and 10 mM $CaCl_2$ (*Right*). (*Upper*) Fluorescein-labeled 3′-amino terminal primer and template DNA duplex with underlined region DNA or NP-DNA.

Following a broader search, we found conditions in which BF is capable of much faster rates of extension beyond a 3′-amino terminus (Fig. 2). We screened divalent metal cofactors under presteady-state conditions and observed strong cofactor-dependent effects on activity. Kinetics of 3′-amino primer extension with 5 mM Ca^{2+}, Mg^{2+}, Ba^{2+}, and Sr^{2+} were all faster than in the absence of added divalent metals (Fig. 2*A*). The cofactor-dependent kinetics did not follow a trend consistent with the Irving–Williams series nor that expected from the aquo ion pK_a. Interestingly, Ca^{2+} had the largest kinetic effect at pH 8.8 and 55 °C. When titrated into the reaction, observed rate constants showed saturation behavior at similar apparent dissociation constants, $K_{d,app}$, for Mg^{2+} and Ca^{2+} (1.3 mM vs. 2.1 mM, respectively, in the presence of 1 mM dCTP). However, the maximum reaction velocity was ∼5-fold faster in the presence of saturating Ca^{2+} vs. Mg^{2+} (Fig. 2 *B* and *C*), maximally ∼4.5 nt/h with Ca^{2+}. Phosphodiester-forming activity with Ca^{2+} has been reported (18, 19). Here, two additions of dCTP were observed on a GG-containing template in the presence of either Ca^{2+} or Mg^{2+}, where the first is linked via a phosphoramidate ester and the second via a phosphodiester (Fig. 2*C*).

With the enhanced kinetics afforded by Ca^{2+}, BF was capable of catalyzing multiple-turnover condensation of nNTP monomers to form NP-DNA oligonucleotides in a DNA template-directed manner (Fig. 2*D*). Extension up to the +8 nt product was detected after 24 h at 55 °C in the presence of Ca^{2+} and all four nNTP substrates, but no significant extension was observed on an NP-DNA template.

Although WT BF has bona fide DNA-dependent NP-DNA polymerase activity, a substantial kinetic defect is associated with nNTP vs. dNTP addition to a 3′-amino terminal primer (Fig. 3*A*). For the first addition to a primer ending in nG, the rate constant, k_{pol}, in reactions with nCTP was ∼22-fold slower than with dCTP (Fig. 3*B* and Table 1). In seeking to rationalize this apparent recognition of the 3′-substituent, we first considered that 3′-amino-2′,3′-dideoxyribonucleosides adopt a C3′-endo sugar conformation and, as such, bear greater conformational similarity to ribonucleosides than deoxyribonucleosides. It is therefore conceivable that nNTPs might be disfavored based on discrimination of their RNA-like sugar pucker. Selectivity against ribonucleotides purportedly involves recognition by "steric gate" residues, and mutations at these positions alter selectivity against ribonucleotide incorporation (20, 21). However, nNTPs and dNTPs are both chemically 2′-deoxy, despite their distinct sugar pseudorotation. Limited screening of mutations at steric gate hotspot positions (Fig. 3*C*), as well as two highly conserved residues involved in metal cofactor binding, mainly exacerbated selectivity against nCTP (Fig. 3*D*). However, mutation at one position in the O helix, Phe-710, known to affect substrate preference for 2′,3′-dideoxyribonucleotide 5′-triphosphates (ddNTPs) (22–24), showed a significant kinetic enhancement, as well as a large effect on dCTP vs. nCTP selectivity. The mutant F710Y demonstrated a 21-fold increase in k_{pol} for nCTP, as well as a 2.6-fold increase for dCTP (Fig. 3*E* and Table 1). This mutant could achieve extension to a full-length +28-nt NP-DNA product in <24 h at 55 °C in the presence of Ca^{2+} and all four nNTPs (Fig. 3*F*).

To establish that NP-DNA polymerase activity is in fact template-directed by BF, we looked for sequence-specific stalling of primer extension in the presence or absence of individual nucleotides. By using dropout mixes, in which one of the four nNTPs is absent in each reaction, we observed a substantial kinetic block at (or proximal to) the position corresponding to the absent substrate (Fig. 3 *G*, *Left*). With drop-in mixes, in which one nucleotide of a dNTP mix is substituted with its corresponding 3′-amino analog in each reaction, BF generates a Sanger-type ladder, with accumulation of products at positions corresponding to the nNTP present in each reaction (Fig. 3 *G*,

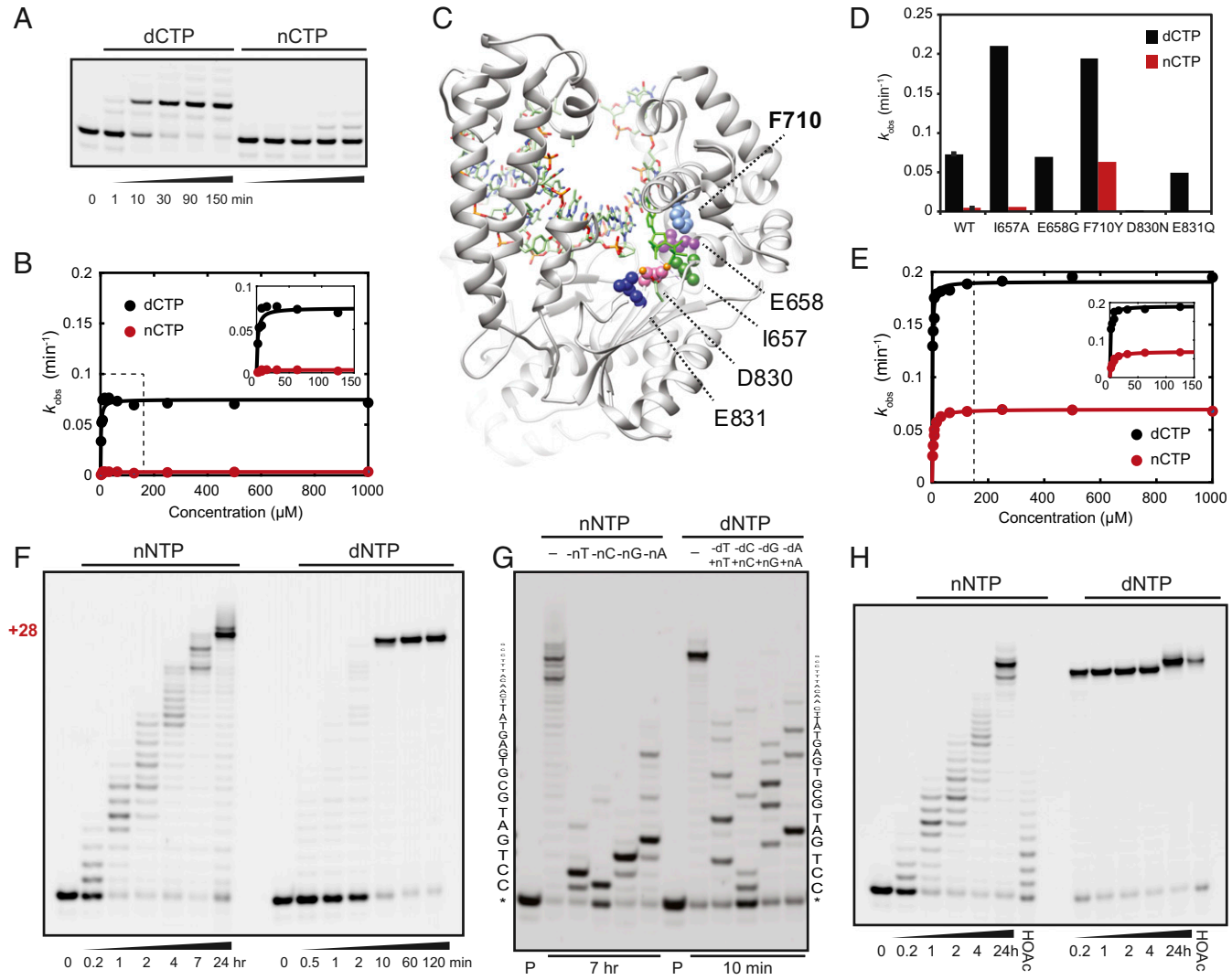

Fig. 3. Discrimination between dNTP and nNTP substrates. (*A*) Time course of 3′-amino primer extension with either 1 mM dCTP or nCTP with wild-type BF and conditions as in Fig. 2. (*B*) Substrate-dependent rate constants for extension of a 3′-amino primer with varying concentrations of dCTP (black) or nCTP (red) with 10 mM $CaCl_2$ under presteady-state conditions. Parameter estimates in Table 1. (*C*) Cartoon of BF closed complex with DNA primer/template and bound substrate (PDB ID code 3EZ5) (39) with mutated active site residues indicated. (*D*) Observed rates of 3′-amino primer extension with either 1 mM dCTP (black) or nCTP (red) for WT BF and various point mutants. (*E*) Substrate-dependent rate constants, as in *B*, for the mutant F710Y (Table 1). (*F*) 3′-Amino primer extension by mutant F710Y on a mixed sequence template (sequence in Fig. 2*D*) with either 1 mM nNTPs or dNTPs at 55 °C in a buffer containing 40 mM Tris·HCl, pH 8.8, 10 mM $CaCl_2$, and 1 mM DTT. (*G*) Extension reactions as in *F* with nNTP drop-out mixes (*Left*: 0.5 mM each nNTP or only three of four, as indicated) or dNTP drop-in mixes (*Right*: 0.5 mM each dNTP or only three of four, as indicated, with fourth complemented with the corresponding nNTP). The gel shows reactions incubated for either 7 h (lanes 2–6) with nNTP dropout mixes or 10 min (lanes 8–12) with dNTP drop-in mixes. "P" indicates lanes containing primer alone. (*H*) 3′-Amino primer extension reactions with F710Y as in *F* but showing acid digest products after 24 h of extension (HOAc lanes).

Right). Finally, the products of these long extension reactions show the acid lability characteristic of NP bonds (Fig. 3*H*). Incubation of the quenched reaction with acetic acid (HOAc) at 75 °C generated a cleavage ladder for all positions when nNTPs were used for 3′-amino primer extension, whereas no hydrolysis intermediates are observed when dNTPs have been used to extend the same primer.

Relative to its native phosphodiester bond forming activity, the kinetic disadvantage of BF's NP-DNA polymerase activity is approximately four orders of magnitude with F710Y, since transient k_{pol}/K_d is 3.1 μM^{-1} s^{-1} for DNA synthesis with Mg^{2+} (25) vs. 3.2 × 10^{-4} μM^{-1} s^{-1} for NP-DNA synthesis with Ca^{2+} (Table 1). This difference is of a similar order to that seen for mismatch incorporation. Unlike mismatches, the binding affinity for cognate nNTPs remains high, and the difference in catalytic efficiency is likely to arise mainly in the chemical step of the reaction. It may nevertheless be instructive to establish whether there are unique structural determinants associated with formation of the phosphoramidate catalytic complex.

To observe the major steps in the reaction pathway for NP-DNA synthesis, from the reactant complex (Figs. 1*B* and 4*A*)

Table 1. Substrate-dependent kinetic parameters for NP activity in BF under presteady-state conditions, as in Fig. 3

	k_{pol}, min^{-1}		$K_{d,app}$, μM		Selectivity*
	dCTP	nCTP	dCTP	nCTP	dCTP/nCTP
WT	0.075	0.0033	1.7	3.8	51
F710Y	0.191	0.069	1.1	3.6	9

*Selectivity defined as the ratio $(k_{pol}/K_{d,app})_{dCTP}/(k_{pol}/K_{d,app})_{nCTP}$.

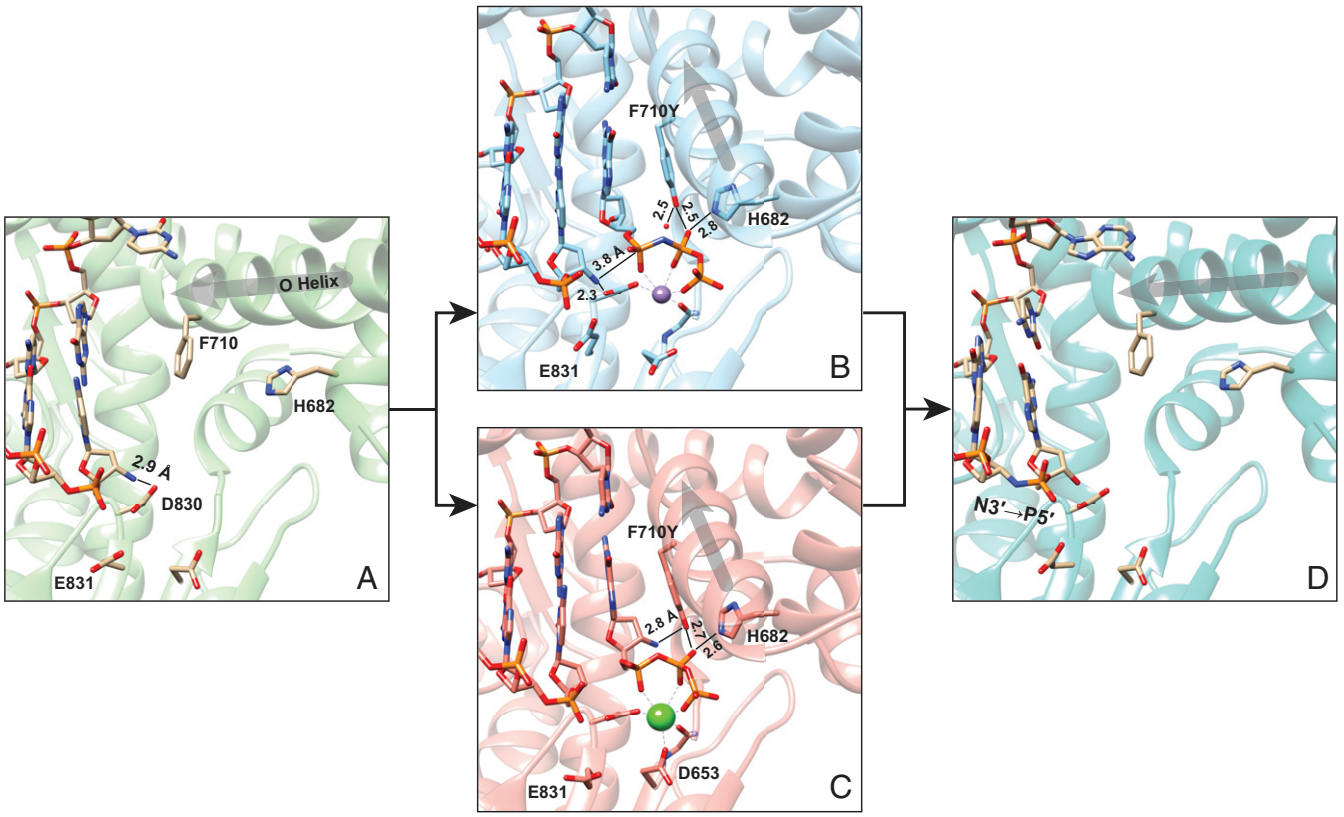

Fig. 4. Structural snapshots of the reaction pathway in NP catalysis. (*A*) Reactant "0 complex" in the open conformation, with bound DNA duplex containing a 3′-amino terminal primer aligned at the active site, as in Fig. 1 *B* and *C* (PDB ID code 6UR4). (*B*) Closed conformation reaction complex bound to DNA duplex containing a 3′-amino terminal primer, substrate analog dGpNHpp, and Mn^{2+} (PDB ID code 6US5). (*C*) Closed conformation reaction complex bound to DNA duplex containing a 3′-dideoxy terminal primer, as well as nGTP substrate and Ca^{2+} (6UR9). (*D*) Posttranslocated product "+1 complex" in the open conformation, containing an incorporated N3′→P5′ bond between the primer terminal nC and dG residues, as in Fig. 1 *D–F* (PDB ID code 6UR2). Atomic distances are indicated in angstroms, and the orientation of the O helix in each structure is shown with a gray arrow.

to the +1 product complex (Figs. 1 *E* and *F* and 4*D*), we cocrystallized BF with an unreactive nucleotide analog (dGpNHpp), a 3′-amino terminated DNA primer, and a DNA template. Using the double mutant F710Y/D598A (24, 26), we crystallized the analog-bound complex with Mn^{2+} and solved the structure to 2.25-Å resolution (Fig. 4*B*). Two complexes were found in the asymmetric unit, both substrates bound in a closed conformation of the fingers domain, as seen by a large conformational change in the O helix proximal to the reaction center. A single metal ion was coordinated by the α,β-imido-triphosphate moiety in each active site, although the bound metal ion and substrate is more ordered in one of the two complexes in the asymmetric unit. In this complex, the terminal 3′-amino group of the primer is in close proximity with the alpha-phosphorus of the analog (N3′-P_α distance ∼3.8 Å), as well as with the carboxylate group of Asp-830 in the active site (N3′-O distance ∼2.3 Å, Fig. 4*C*), significantly closer than the ∼2.9–3.0 Å seen in the open conformation (Fig. 4*A*). This highly conserved aspartate coordinates divalent metal ions in structures of the equivalent native complex when the primer bears a 3′-OH, but there is substantial debate over its potential role as a general base in native phosphodiester catalysis (27–29). In the closed substrate-bound structure, the aspartate conformation is not significantly altered as a result of 3′-amino substitution at the primer terminus relative to that seen previously with 2′,3′-dideoxy termini (26). Nevertheless, this carboxylate appears to be critical, as the mutation D830N entirely abolishes NP-DNA synthesis (Fig. 3*D*). However, the mutation E831Q did not substantially affect activity, although this adjacent residue is typically found coordinating a metal ion in structures modeling the native DNA polymerase activity. In the analogous nonenzymatic chemistry with phosphorimidazolides, rapid kinetics are observed in the absence of divalent metal ions, suggesting that inner sphere metal coordination is superfluous for the reactivity of the 3′-amino nucleophile (6, 7, 30).

An alternative model of the catalytic complex was produced by cocrystallization with nGTP and a template-bound primer carrying a 3′-terminal 2′,3′-dideoxycytidine (ddC) residue. This complex with the BF double-mutant was crystallized and solved to 2.10-Å resolution (Fig. 4*C*). The resulting structure shows a single ordered Ca^{2+} and substrate in the closed conformation, but only in one of two complexes in the asymmetric unit (Fig. 5*A*). In the closed complex, a possible role for the mutation F710Y is suggested by the orientation of the tyrosine hydroxyl, which is in close proximity to both the β-phosphate nonbridging oxygen of the incoming substrate and its 3′-amino group. This position is similar to that seen in the presence of a dideoxynucleotide substrate (24). The noncovalent bonding network also appears to involve the heterocyclic nitrogen of His-682 (Fig. 4 *B* and *C*). Moderate donor-acceptor distances across this network, 2.6–2.8 Å, suggest a distributed hydrogen-bonding stabilization of the substrate and leaving group (Fig. 5*A*). In the dGpNHpp analog-bound reaction complex, the C2′-endo sugar conformation of the substrate analog is such that a solvent water is bound within this network (Fig. 4*B*), whereas the 3′-amino group of nGTP in the C3′-endo conformation appears to substitute for this water in the alternative closed complex (Fig. 4*C*). However, the apparent binding affinity for nCTP vs. dCTP was unaffected by

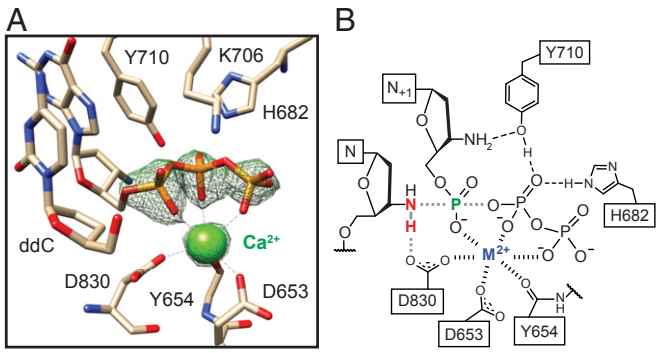

Fig. 5. Structural model for the N-P bond forming reaction in BF. (*A*) Structure and $F_o - F_c$ omit map (green mesh, 5 σ) for the observed closed reaction complex bound to DNA duplex containing a ddC terminal primer, nGTP substrate, and Ca^{2+}. (*B*) Model of the reaction complex with attacking 3′-amino group (red) of the primer terminus (N) in reaction with the α-phosphate center (green) of an incoming substrate (N_{+1}) and a coordinated metal ion (M^{2+}, blue). Interacting residues found in closed complex structures are indicated.

F710Y with Ca^{2+}, indicating that this tuning of the substrate is likely relevant only to its conformation in the closed catalytic complex.

Discussion

More-efficient NP catalysis is likely to be attainable by directed evolution of BF, as suggested by the observation that 3′-amino extension remained selective for dCTP over nCTP even in the improved point mutant (Fig. 3 and Table 1). This selectivity is consistent with the observation that the bound substrate geometry of nGTP seen in the crystal structure is distinct from that observed for the unreactive dGTP analog (Fig. 4 *B* and *C*). Tuning of the reactivity of the substrate appears to be conferred not only by triphosphate-metal coordination, but also by side chains in the fingers domain that interact directly with the bound substrate in the closed conformation.

Phosphodiester bonds form the genetic backbone of life on Earth, but phosphoramidate esters do not. It should therefore be unsurprising that highly evolved polymerases are optimized for their WT activity and substrate, but it remains notable that 3′-amino nucleotides are not strictly chain terminators. Extension beyond a 3′-amino terminus proceeds with an alternative metal cofactor preference, suggesting mechanistic distinctions in the chemical step between N-P and O-P bond formation that might be further optimized. Since the pK_a of the protonated 3′-amino group is ∼7.5–7.7 as the free nucleoside (31, 32), consistent with amino functionalities in other glycosylamines (33), the amino nucleophile is expected to be substantially neutral under the reaction conditions. Notably, mutation of two acidic active site residues, which canonically bind a metal ion, activating the nucleophilic 3′-OH in the native O-P reaction, had disparate effects on the N-P reaction. The mutation D830N completely eliminated 3′-amino primer extension, but E831Q did not (Fig. 3*D*). The kinetics of these mutants, in addition to structural evidence, are more consistent with a role for Asp-830 as a general base than as a metal ligand proximal to the nucleophile (Fig. 5*B*). We therefore expect that this aspartate is likely to play a key role in facilitating proton transfer out of the transition state, in this case without the inner sphere nucleophile-metal ion coordination that is widely understood to activate the nucleophile in the corresponding phosphodiester forming reaction (28). It is therefore plausible that the distinction in proton transfer between the two reaction mechanisms contributes substantially to the observed kinetic defect. Some additional outer sphere role for divalent ions in the mechanism also cannot be ruled out, since presteady-state rates were not saturated at one equivalent of metal ion vs. substrate (Fig. 2). Ca^{2+}-mediated catalysis of DNA polymerase activity, which we incidentally observed here, has been previously reported (18, 19), but the precise role of Ca^{2+} in the phosphodiester-forming mechanism remains to be elucidated. By cocrystallization, we observe Ca^{2+} bound to the substrate triphosphate moiety in a manner similar to that seen for Mg^{2+} crystallized under similar conditions, but there remains the possibility that our structures do not capture transiently bound metal ions that are nevertheless critical to the reaction mechanism, as has recently been proposed for certain eukaryotic polymerase family members (34, 35).

For NP-DNA to participate in any abiogenesis, e.g., a synthetic one undertaken in the laboratory, it would require that this alternative genetic polymer play informational and functional roles in the resulting cell. Given that DNA, RNA, and various xenonucleic acids (XNAs) can form folded structures with a broad range of catalytic activities stretching across sequence space, it is reasonable to expect that the sequence space of NP-DNA, too, will contain diverse functions. Although several key aspects of its viability as an alternative genetic polymer remain to be explored, it has long been known that NP-DNA forms stable duplexes with RNA and DNA (36), implying its capacity to exchange information. Among alternative nucleic acids known to pair with RNA, only a subset (37) are compatible with template-directed nonenzymatic copying chemistry based on phosphorimidazolides. Interestingly, NP-DNA now appears to be compatible with both nonenzymatic phosphorimidazolide and enzymatic triphosphate chemistries. As a result, a genetic transition between backbone linkages is biochemically conceivable with either chemistry.

Materials and Methods

Materials. 2′-Deoxyribonucleotide 5′-triphosphates were obtained from New England Biolabs. 3′-Amino-2′,3′-dideoxyribonucleotide 5′-triphosphates were obtained from Trilink Biotechnologies. DNA oligonucleotides were purchased from IDT. Oligonucleotides incorporating a 3′-amino residue were synthesized and purified by standard methods, using phosphoramidites obtained from Chemgenes or as recently reported (13) and using reagents purchased from Glen Research on an Expedite 8909 synthesizer. Premixed reagent solutions for crystallization were purchased from Hampton Research. Recombinant lysozyme and Bugbuster lysis reagent were purchased from EMD. Ni-NTA Superflow affinity resin was purchased from Qiagen. All other buffers and reagents were prepared from high-purity chemicals obtained from Sigma-Aldrich, using chemicals with trace metals analysis where available. Competent cells were obtained from New England Biolabs.

Recombinant Protein Preparation. A bacterial expression plasmid, carrying the codon-optimized sequence for the DNA polymerase I large fragment (amino acids 297–877) from Bst (BF) fused at its N terminus to a His_6 tag and an HRV 3C protease signal sequence, was produced by standard methods. BF expression was placed under the control of a lacO-regulated T5 promoter based on the expression system from pQE-30. The plasmid backbone carries a kanamycin resistance cassette, replication origin, and lacI derived from pET-28. The resulting plasmid, pVSL5, was used to transform DH5α cells selected on LB plates containing kanamycin. For protein expression, starter cultures were inoculated from a single colony and grown overnight in LB + 1% glucose and kanamycin at 30 μg/mL at 30 °C. The overnight culture was then inoculated 1:100 into a shake flask containing LB and antibiotic, and the culture was expanded at 37 °C with 225 rpm shaking (e.g., with 0.5 L of medium in a 2-L flask). When the optical density at 600 nm reached ∼0.6–0.8 in a 1-cm path length cuvette, expression of BF was induced by the addition of 1 mM IPTG, and induction was allowed to proceed for 2 h at 37 °C. Cells were harvested by centrifugation at 4 °C, and the resulting cell pellet was typically stored at −80 °C prior to purification. Thawed pellets were lysed in 1× BugBuster, adjusted to pH ∼8.5 by addition of KOH, containing 0.001× ethylenediaminetetraacetic acid (EDTA)-free Protease Inhibitor Mixture Set III (Calbiochem) and lysozyme. The lysate was heat treated in a water bath at 50 °C for 20 min, and its viscosity was reduced by probe sonication at 0 °C. The lysate was then clarified by centrifugation at 16,000 x *g* for 30 min at 4 °C. The polyhistidine-tagged protein was purified by batch binding to Ni-NTA beads preequilibrated in 1× BugBuster. The resin was washed with 1 column volume (CV) of BugBuster, followed by 10 CV of 20 mM sodium phosphate, pH 7.6, 0.5 M NaCl, 12.5 mM imidazole. An additional wash with a higher salt buffer, otherwise as above except containing 1 M NaCl, reduced nonspecific binding. The protein was eluted with 20 mM sodium phosphate, pH 7.6, 0.5 M NaCl, 250 mM imidazole,

and dithiothreitol (DTT) was added to the resulting fractions up to 2 mM. The protein was concentrated using centrifugal 50 kDa molecular weight cut-off (MWCO) filtration devices (Millipore), desalted into a 2× storage buffer using a disposable size exclusion column (NAP25, GE Amersham) following the manufacturer's protocol, and then reconcentrated. The 2× storage buffer contained 20 mM Tris-Cl, pH 7.6, 100 mM KCl, 2 mM DTT, 0.2 mM EDTA, and 0.2% Triton X-100. The resulting protein preparation was diluted twofold by gradual addition of 1 volume of glycerol, then stored at −30 °C. To prepare protein for crystallization, elution fractions from the Ni-NTA resin were diluted 1:1 by the addition of the 2× storage buffer without Triton X-100 and concentrated, followed by HRV 3C protease for cleavage of the affinity tag. The cleavage reaction proceeded overnight at 4 °C, and the His-tagged protease and cleaved peptide were removed by passing the solution over Ni-NTA resin, followed by concentration and desalting into 1× storage buffer, as above except omitting the nonionic surfactant.

Cocrystallizations. Preparations of the WT or D598A/F710Y double-mutant BF were concentrated to 20 mg/mL. The mutation D598A reportedly disrupts a crystal contact that otherwise precludes crystallization of the closed conformation of the fingers domain (26). A DNA primer, containing either a 3′ terminal 2′,3′-dideoxycytidine (5′-GCGATCAGddC) or 3′-amino-2′,3′-dideoxycytidine (5′-GCGATCAGnC) residue, was prepared and annealed with a partially complementary DNA template oligonucleotide (5′-ACACGCTGATCGCA) in the presence of either dGTP, nGTP, or dGpNHpp, as indicated. The final concentrations of primer/template duplex and substrate were 0.25 mM and 3 mM, respectively. Protein was added to a final concentration of 10 mg/mL and allowed to complex by incubating at 37 °C for ∼10–20 min or overnight at 4 °C. To obtain crystals of the substrate-bound closed conformation, mixtures of the D598A/F710Y double mutant protein containing substrate, primer, and template were screened by mixing 1:1 with crystallization reagent solutions containing 0.1 M Na-MES pH 5.4 or 5.8 buffer, 2 M ammonium sulfate, 2.5% or 5% (vol/vol) (±)-2-methyl-2,4-pentanediol (MPD), and 10 mM of $MgSO_4$, $CaCl_2$, $MnSO_4$, or $CoCl_2$. Structures in the closed conformation with ordered substrate and metal ion were obtained only with calcium, manganese, or cobalt ions. Crystallization proceeded by the sitting-drop vapor diffusion method. Crystals typically appeared within 2–4 d at 18 °C. Prior to mounting and freezing, crystals were soaked in a solution of the same reagent mix as for crystallization, except where the buffer was replaced with 0.1 M Na-EPPS, pH 8.8, and the substrate concentration was 0.5 mM. Soak time was screened from ∼2 min to 90 min at 20 °C. For the closed conformation structure containing Ca^{2+} and dGpNHpp (6UR9), the crystal was formed at pH 5.4 with 2.5% MPD, and the soak time at pH 8.8 was ∼20 min. For the closed conformation structure containing Mn^{2+} and nGTP (6US5), the crystal was formed at pH 5.4 with 5% MPD, and the soak time was ∼5 min at pH 8.8. Crystals of the WT BF complex in the open conformation were prepared similarly, except using 8 mM dGTP and 20 mM $MgCl_2$ in the protein complex mix, and these crystals were mounted and frozen directly from the mother liquor without any additional soaking. No ordered density corresponding dGTP or Mg^{2+} were observed in these open conformation WT structures. The prereaction open structure (0 Complex, 6UR4) was crystallized by mixing the protein complex solution 1:1 with 60% vol/vol Tacsimate, pH 7.0. Tascimate (Hampton Research) is a reagent mix containing 1 M malonic acid, 0.15 M ammonium citrate tribasic, 0.07 M Succinic acid, 0.18 M DL-malic acid, 0.24 M sodium acetate trihydrate, 0.3 M sodium formate, 0.1 M ammonium tartrate dibasic, pH 7.0. The postreaction translocated open structure (+1 Complex, 6UR2), with incorporated dGTP, was obtained after ∼45 d by mixing the same protein complex solution 1:1 with 0.8 M sodium succinate, pH 7.0. For all crystals, the space group was $P2_12_12_1$, and the structures were solved by molecular replacement.

Primer Extension Assays. Except where indicated, all extension reactions were performed with 1 μM fluorescein-labeled primer, 1.5 μM template, and ∼1.1 μM purified His_6-tagged protein at 55 °C in a reaction buffer composed of 40 mM Tris·HCl, pH 8.8, 1 mM DTT, and a divalent cation as specified, typically 10 mM $CaCl_2$. Reaction mixtures were equilibrated at the indicated reaction temperature for 1 min and initiated by addition of substrate. Progress was monitored by sampling 1 μL from the reaction manually quenched into 24 μL of chilled 90% formamide, 10 mM EDTA. Quenched samples were denatured at 90 °C for 1 min and cooled to room temperature prior to separation by denaturing polyacrylamide gel electrophoresis on 20% Tris-borate-EDTA (TBE) urea gels. To maintain strand separation for long extension products, as in Fig. 3 *F–H*, an unlabeled DNA oligonucleotide complementary to the template was added to the quenching buffer (∼2 μM) prior to denaturation. Typically, 3 or 4 μL of quenched sample was loaded per lane for gels of 0.75-mm thickness. Bands were visualized with a Typhoon laser scanning imager (GE Amersham) and quantified with the manufacturer's software. For acid digestion of extended products, a reaction sample was diluted with 2.5 volumes of 1% acetic acid, 10 mM EDTA in formamide and digested at 75 °C for 45 min prior to gel separation.

Mass Spectrometry. Reaction samples were desalted using C18 ZipTips equilibrated with liquid chromatography mass spectrometry (LC-MS) grade methanol followed by 2 M triethylammonium acetate, pH 7 (TEAA). The sample was diluted by addition of TEAA buffer to a concentration of 0.2 M and bound to the support. The resin was then washed with at least 100 μL of 20 mM TEAA. The sample was eluted in 50% methanol directly into an HPLC vial insert and dried in a centrifugal vacuum concentrator. Samples were resuspended in LC-MS grade water prior to injection. LC-MS separation and analysis were performed essentially as described previously (38).

Data Availability Statement. Structural data are available via the Protein Data Bank (PDB), http://www.wwpdb.org/, under ID codes 6UR2 (40), 6UR4 (41), 6UR9 (42), and 6US5 (43). Plasmids are available upon request.

ACKNOWLEDGMENTS. J.W.S. is an Investigator of the Howard Hughes Medical Institute. This work was funded in part by Simons Foundation Grant 290363 (to J.W.S.).

1. A. Eschenmoser, M. V. Kisakürek, Chemistry and the origin of life. *Helv. Chim. Acta* **79**, 1249–1259 (1996).
2. L. E. Orgel, *The Origins of Life: Molecules and Natural Selection* (Wiley, New York, 1973), (September 1, 2019).
3. W. S. Zielinski, L. E. Orgel, Autocatalytic synthesis of a tetranucleotide analogue. *Nature* **327**, 346–347 (1987).
4. A. Eschenmoser, Chemical etiology of nucleic acid structure. *Science* **284**, 2118–2124 (1999).
5. D. K. O'Flaherty, L. Zhou, J. W. Szostak, Nonenzymatic template-directed synthesis of mixed-sequence 3′-NP-DNA up to 25 nucleotides long inside model protocells. *J. Am. Chem. Soc.* **141**, 10481–10488 (2019).
6. S. Zhang, N. Zhang, J. C. Blain, J. W. Szostak, Synthesis of N3′-P5′-linked phosphoramidate DNA by nonenzymatic template-directed primer extension. *J. Am. Chem. Soc.* **135**, 924–932 (2013).
7. S. Zhang, J. C. Blain, D. Zielinska, S. M. Gryaznov, J. W. Szostak, Fast and accurate nonenzymatic copying of an RNA-like synthetic genetic polymer. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17732–17737 (2013).
8. R. L. Letsinger, J. S. Wilkes, L. B. Dumas, Enzymatic synthesis of polydeoxyribonucleotides possessing internucleotide phosphoramidate bonds. *J. Am. Chem. Soc.* **94**, 292–293 (1972).
9. Z. G. Chidgeavadze *et al*., 2′,3′-Dideoxy-3′ aminonucleoside 5′-triphosphates are the terminators of DNA synthesis catalyzed by DNA polymerases. *Nucleic Acids Res.* **12**, 1671–1686 (1984).
10. Y. C. Cheng, G. E. Dutschman, K. F. Bastow, M. G. Sarngadharan, R. Y. Ting, Human immunodeficiency virus reverse transcriptase. General properties and its interactions with nucleoside triphosphate analogs. *J. Biol. Chem.* **262**, 2187–2189 (1987).
11. A. L. Handlon, N. J. Oppenheimer, Thiol reduction of 3′-azidothymidine to 3′-aminothymidine: Kinetics and biomedical implications. *Pharm. Res.* **5**, 297–299 (1988).
12. S. Shuman, D. N. A. Ligases, DNA ligases: Progress and prospects. *J. Biol. Chem.* **284**, 17365–17369 (2009).
13. V. S. Lelyveld, D. K. O'Flaherty, L. Zhou, E. C. Izgu, J. W. Szostak, DNA polymerase activity on synthetic N3′→P5′ phosphoramidate DNA templates. *Nucleic Acids Res.* **47**, 8941–8949 (2019).
14. J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature* **391**, 304–307 (1998).
15. D. Ding *et al*., An oligodeoxyribonucleotide N3′–> P5′ phosphoramidate duplex forms an A-type helix in solution. *Nucleic Acids Res.* **24**, 354–360 (1996).
16. D. Ding, S. M. Gryaznov, W. D. Wilson, NMR solution structure of the N3′ –> P5′ phosphoramidate duplex d(CGCGAATTCGCG)2 by the iterative relaxation matrix approach. *Biochemistry* **37**, 12082–12093 (1998).
17. V. Tereshko, S. Gryaznov, M. Egli, Consequences of replacing the DNA 3′-oxygen by an amino group: High-resolution crystal structure of a fully modified N3′ → P5′ phosphoramidate DNA Dodecamer duplex. *J. Am. Chem. Soc.* **120**, 269–283 (1998).
18. A. Irimia *et al*., Calcium is a cofactor of polymerization but inhibits pyrophosphorolysis by the Sulfolobus solfataricus DNA polymerase Dpo4. *Biochemistry* **45**, 5949–5956 (2006).
19. C. Ralec, E. Henry, M. Lemor, T. Killelea, G. Henneke, Calcium-driven DNA synthesis by a high-fidelity DNA polymerase. *Nucleic Acids Res.* **45**, 12425–12440 (2017).
20. C. M. Joyce, Choosing the right sugar: How polymerases select a nucleotide substrate. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1619–1622 (1997).
21. J. A. Brown, Z. Suo, Unlocking the sugar "steric gate" of DNA polymerases. *Biochemistry* **50**, 1135–1142 (2011).
22. S. Tabor, C. C. Richardson, A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6339–6343 (1995).

23. M. Astatke, N. D. F. Grindley, C. M. Joyce, How E. coli DNA polymerase I (Klenow fragment) distinguishes between deoxy- and dideoxynucleotides. *J. Mol. Biol.* **278**, 147–165 (1998).
24. J. J. Warren, L. J. Forsberg, L. S. Beese, The structural basis for the mutagenicity of O(6)-methyl-guanine lesions. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19701–19706 (2006).
25. W. Wang, E. Y. Wu, H. W. Hellinga, L. S. Beese, Structural factors that determine selectivity of a high fidelity DNA polymerase for deoxy-, dideoxy-, and ribonucleotides. *J. Biol. Chem.* **287**, 28215–28226 (2012).
26. S. J. Johnson, J. S. Taylor, L. S. Beese, Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3895–3900 (2003).
27. C. Castro *et al*., Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4267–4272 (2007).
28. T. A. Steitz, S. J. Smerdon, J. Jäger, C. M. Joyce, H. Pelletier, A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* **266**, 2022–2025 (1994).
29. H. Pelletier, M. R. Sawaya, A. Kumar, S. H. Wilson, J. Kraut, Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science* **264**, 1891–1903 (1994).
30. M. Röthlingshöfer, C. Richert, Chemical primer extension at submillimolar concentration of deoxynucleotides. *J. Org. Chem.* **75**, 3945–3952 (2010).
31. E. C. Izgu, S. S. Oh, J. W. Szostak, Synthesis of activated 3′-amino-3′-deoxy-2-thiothymidine, a superior substrate for the nonenzymatic copying of nucleic acid templates. *Chem. Commun. (Camb.)* **52**, 3684–3686 (2016).
32. E. Kervio, A. Hochgesand, U. E. Steiner, C. Richert, Templating efficiency of naked DNA. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12074–12079 (2010).
33. S. Inoue, On the prediction of pKa values of amino sugars. *Chem. Pharm. Bull. (Tokyo)* **16**, 1134–1137 (1968).
34. Y. Gao, W. Yang, Capture of a third Mg^{2+} is essential for catalyzing DNA synthesis. *Science* **352**, 1334–1337 (2016).
35. B. D. Freudenthal *et al*., Uncovering the polymerase-induced cytotoxicity of an oxidized nucleotide. *Nature* **517**, 635–639 (2015).
36. S. Gryaznov, J.-K. Chen, Oligodeoxyribonucleotide N3′->P5′ phosphoramidates: Synthesis and Hybridization properties. *J. Am. Chem. Soc.* **116**, 3143–3144 (1994).
37. S. C. Kim *et al*., A model for the emergence of RNA from a prebiotically plausible mixture of ribonucleotides, arabinonucleotides and 2′-deoxynucleotides. bioRxiv: 10.1101/813675 (2 November 2019).
38. A. Björkbom *et al*., Bidirectional Direct sequencing of Noncanonical RNA by two-Dimensional analysis of mass Chromatograms. *J. Am. Chem. Soc.* **137**, 14430–14438 (2015).
39. A. A. Golosov, J. J. Warren, L. S. Beese, M. Karplus, The mechanism of the translocation step in DNA replication by DNA polymerase I: A computer simulation analysis. *Structure* **18**, 83–93 (2010).
40. V. S. Lelyveld, W. Zhang, J. W. Szostak, DNA polymerase I Large Fragment from Bacillus stearothermophilus with DNA template and primer containing an N3′→P5′ linkage. Protein Data Bank. Available at https://www.rcsb.org/structure/6UR2. Deposited 21 November 2019.
41. V. S. Lelyveld, W. Zhang, J. W. Szostak, DNA polymerase I Large Fragment from Bacillus stearothermophilus with DNA template and 3′-amino primer. Protein Data Bank. Available at https://www.rcsb.org/structure/6UR4. Deposited 22 November 2019.
42. V. S. Lelyveld, W. Zhang, J. W. Szostak, DNA polymerase I Large Fragment from Bacillus stearothermophilus with DNA template, dideoxy primer, 3′-amino-ddGTP (nGTP), and Ca^{2+}. Protein Data Bank. Available at https://www.rcsb.org/structure/6UR9. Deposited 22 October 2019.
43. V. S. Lelyveld, W. Zhang, J. W. Szostak, DNA polymerase I Large Fragment from Bacillus stearothermophilus with DNA template, 3′-amino primer, dGpNHpp analog, and Mn^{2+}. Protein Data Bank. Available at https://www.rcsb.org/structure/6US5. Deposited 24 October 2019.