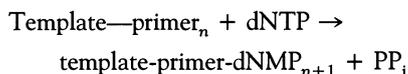


# Structures of Ternary Complexes of Rat DNA Polymerase $\beta$ , a DNA Template-Primer, and ddCTP

Huguette Pelletier, Michael R. Sawaya, Amalendra Kumar, Samuel H. Wilson, Joseph Kraut

Two ternary complexes of rat DNA polymerase  $\beta$  (pol  $\beta$ ), a DNA template-primer, and dideoxycytidine triphosphate (ddCTP) have been determined at 2.9 Å and 3.6 Å resolution, respectively. ddCTP is the triphosphate of dideoxycytidine (ddC), a nucleoside analog that targets the reverse transcriptase of human immunodeficiency virus (HIV) and is at present used to treat AIDS. Although crystals of the two complexes belong to different space groups, the structures are similar, suggesting that the polymerase-DNA-ddCTP interactions are not affected by crystal packing forces. In the pol  $\beta$  active site, the attacking 3'-OH of the elongating primer, the ddCTP phosphates, and two  $Mg^{2+}$  ions are all clustered around Asp<sup>190</sup>, Asp<sup>192</sup>, and Asp<sup>256</sup>. Two of these residues, Asp<sup>190</sup> and Asp<sup>256</sup>, are present in the amino acid sequences of all polymerases so far studied and are also spatially similar in the four polymerases—the Klenow fragment of *Escherichia coli* DNA polymerase I, HIV-1 reverse transcriptase, T7 RNA polymerase, and rat DNA pol  $\beta$ —whose crystal structures are now known. A two-metal ion mechanism is described for the nucleotidyl transfer reaction and may apply to all polymerases. In the ternary complex structures analyzed, pol  $\beta$  binds to the DNA template-primer in a different manner from that recently proposed for other polymerase-DNA models.

DNA replication (1) is a highly complex biological process, even for a relatively simple organism such as *Escherichia coli*. During replication, the double helical DNA molecule is unwound, and the two resultant single strands of DNA act as templates to guide the synthesis, one complementary base at a time, of antiparallel primer strands. Although many auxiliary proteins such as ligases, helicases, and topoisomerases are usually involved, the chemical reaction at the core of DNA replication, the nucleotidyl transfer reaction, is catalyzed by DNA polymerases and may be depicted as follows:



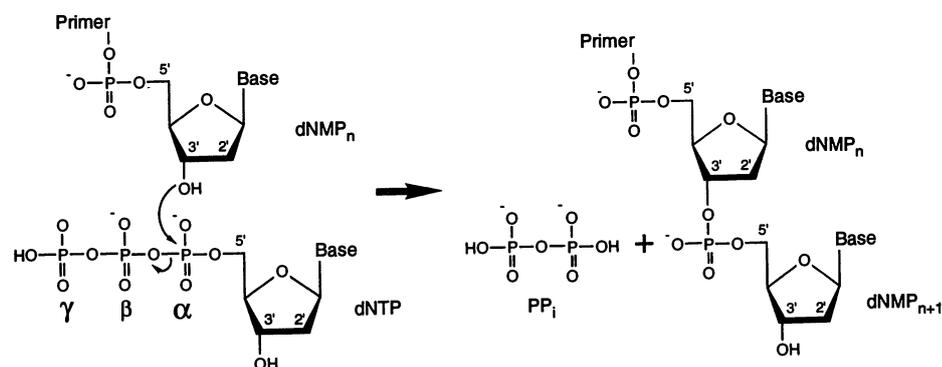
where dNTP (2'-deoxyribonucleoside 5'-triphosphate) represents any one of four deoxynucleotides (dATP, dGTP, dCTP, and dTTP), and dNMP and  $\text{PP}_i$  represent 2'-deoxyribonucleoside 5'-monophosphate and pyrophosphate, respectively (Fig. 1).

Inhibition of a polymerase that effects genomic replication can be fatal to an organism. In a common type of polymerase inhibition, 2',3'-dideoxynucleotides (ddNTPs) act as chain terminators of the primer strand. The ddNTPs differ from

their cellular dNTP counterparts by the absence of an attacking 3'-hydroxyl group (3'-OH) (Fig. 2) and therefore, once a dideoxynucleotide is successfully incorporated into a growing primer strand, there can be no further incorporation of subsequent nucleotides. A well-known example of this kind of inhibition involves HIV-1 reverse transcriptase (RT), which is the polymerase responsible for the replication of the HIV genome. 3'-azido-2',3'-dideoxythymidine (AZT), 2',3'-dideoxyinosine

(ddI), and 2',3'-dideoxycytidine (ddC) are all anti-HIV drugs (2, 3) that become potent chain termination inhibitors of RT after they are converted by cellular kinases (4, 5), in vivo, to their corresponding nucleoside 5'-triphosphates, AZT-TP, ddATP (6), and ddCTP, respectively. In that all polymerases probably share a common catalytic mechanism, it is not surprising that some toxic effects of these drugs have been attributed to inhibition of host-cell polymerases, perhaps including the pol  $\beta$  described here (7–9). Therefore, a detailed understanding of the nucleotidyl transfer reaction, as well as the mechanism of inhibition of viral and host cell polymerases by nucleoside analogs, may lead to the design of more potent and less toxic HIV-1 RT inhibitors for use in the treatment of AIDS.

Despite limited sequence similarity to the Klenow fragment (KF) of *E. coli* DNA pol I [the only other polymerase for which a crystal structure (10) was known at the time], the crystal structure determinations of HIV-1 RT (11, 12) revealed a common polymerase fold consisting of three distinct subdomains (designated fingers, palm, and thumb because of the resemblance to a hand) forming an obvious DNA binding channel. The strongest structural overlap between KF and RT comprised a trio of carboxylic acid residues located in the palm subdomain (11, 12). These observations led to the hypothesis that perhaps all polymerases share a common nucleotidyl transfer mechanism centered around the highly conserved carboxylic acid residues (11). Strengthening this argument somewhat was the subsequent crystal structure determination of an RNA polymerase (RNAP) from bacteriophage T7, which showed strong structural similarities with KF (13). How-

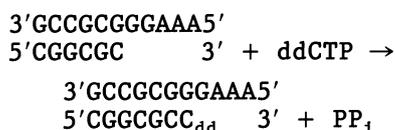


**Fig. 1.** The nucleotidyl transfer reaction. The 3'-OH group of the terminal dNMP on the primer strand attacks the 5'- $\alpha$  phosphate of an incoming dNTP, and a newly formed phosphodiester linkage results in elongation of the primer strand by one dNMP. After release of pyrophosphate ( $\text{PP}_i$ ), the catalytic cycle is complete and the 3'-OH group of the newly incorporated dNMP is now ready to attack yet another incoming dNTP. Only the 3' end of a primer is extended so that DNA polymerization is said to proceed in a 5' to 3' direction. If the polymerase molecule does not release the template-primer before incorporation of a second dNMP, the mode of DNA synthesis is said to be "processive", but if the polymerase releases the template-primer after each successive incorporation of a dNMP, the mode of DNA synthesis is said to be "distributive".

H. Pelletier, M. R. Sawaya, and J. Kraut are in the Department of Chemistry, University of California, San Diego, CA 92093-0317, USA. A. Kumar and S. H. Wilson are at the Sealy Center for Molecular Science, University of Texas Medical Branch, Galveston, TX 77555-1051, USA.

ever, because structural evidence for a common nucleotidyl transfer mechanism has so far been limited to comparisons among polymerases from a bacterium (KF), a virus (RT), and a phage (RNAP), perhaps the most convincing evidence for this hypothesis is provided by the crystal structure determination of a eukaryotic polymerase, rat DNA pol  $\beta$  (14). Sequence alignments show that pol  $\beta$  is so distantly related, even from its eukaryotic relatives, polymerases  $\alpha$ ,  $\gamma$ ,  $\delta$ , and  $\epsilon$ , that it stands in a class of its own along with only one other polymerase, terminal deoxynucleotidyltransferase (TdT) (15). The crystal structure of pol  $\beta$  nevertheless revealed a polymerase fold consisting of palm, fingers, and thumb (along with an additional 8-kD domain attached to the fingers), and the most striking structural similarity with its distant relatives, KF, RT, and RNAP, is a portion of the palm that bears the highly conserved carboxylic acid residues (14). This suggests that despite the large differences in size (pol  $\beta$ , at 39 kD, is the smallest polymerase known), in function [although pol  $\beta$  may play a role in DNA replication (16, 17), its primary function is in DNA repair (18–20)], and in fidelity [pol  $\beta$  is the most error prone eukaryotic polymerase studied to date (21, 22)], pol  $\beta$  probably shares a common nucleotidyl transfer catalytic mechanism with all other polymerases.

Taking advantage of the chain termination method of polymerase inhibition with ddNTPs, we have succeeded in growing crystals of rat pol  $\beta$  complexed with two pseudo substrates, namely, (i) a DNA template-primer in which the 3' end of the primer has been "terminated" by ddCMP, and (ii) ddCTP. In preparation for crystallization experiments, we mixed pol  $\beta$  with the DNA template-primer shown below and a large excess of ddCTP on the assumption that, prior to crystallization, the following reaction would occur:



where  $C_{\text{dd}}$  is the newly incorporated ddCMP. If pol  $\beta$  then were to try to incorporate another nucleotide onto the primer terminus, a second nucleotidyl transfer reaction could not occur because the recently incorporated ddCMP lacks a 3'-OH group. This should result in a pseudo Michaelis-Menten ternary complex in which both "substrates" are present, namely, a nonreactive template-primer and a nucleoside triphosphate. Crystals were obtained, and the subsequent structure determinations revealed that this must have been what happened. Electron density maps

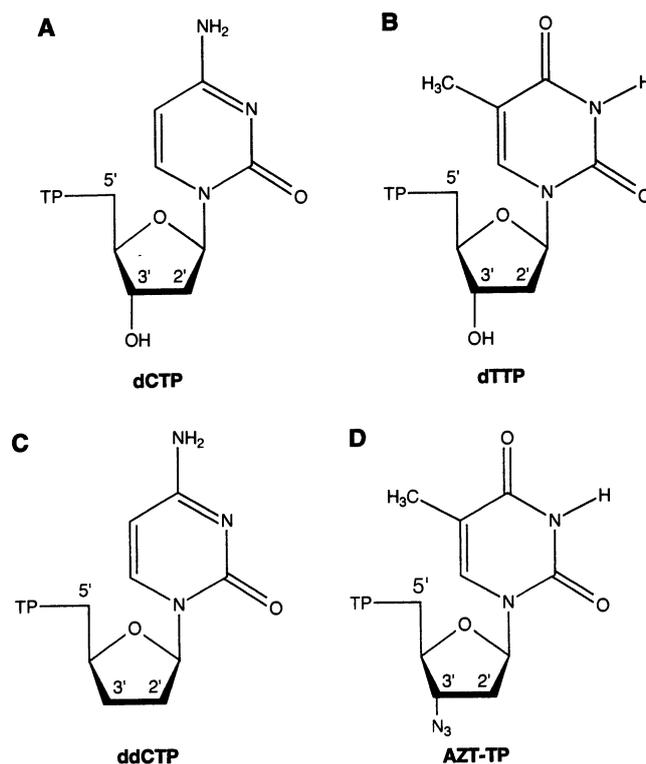
showed a primer strand that was seven nucleotides long, although we started with a primer that was only six nucleotides in length, and the 3' (deoxy) terminus of the primer was positioned next to strong electron density resembling a nucleoside triphosphate, presumably ddCTP.

Such a detailed view of the active site in the ternary complex allows us to propose a two-metal ion mechanism for the nucleotidyl transfer reaction that is similar, in many ways, to the two-metal ion mechanism previously proposed for another type of phosphoryl transfer reaction—the exonuclease reaction of the 3'→5' exonuclease of *E. coli* DNA pol I (23, 24). Our proposed nucleotidyl transfer mechanism probably applies to all polymerases, but when we attempt to extend that mechanism to the other three polymerases—KF, RT, and RNAP—for which the crystal structures are known, a problem arises: in our structures, pol  $\beta$  is bound to the DNA in a manner that differs from the recently proposed polymerase-DNA models for all three of these polymerases.

**Crystallizations and preliminary diffraction studies.** Recombinant rat DNA pol  $\beta$  (25) was expressed in *E. coli* and purified as described (26). After purification, the protein was washed three times in a microconcentrator (Centricon-10, Amicon) with a buffer solution (10 mM ammonium sulfate, 0.1 M tris, pH 7.0), then concentrated to 20 mg/ml and stored at  $-80^\circ\text{C}$  in sealed Eppendorf tubes (120- $\mu\text{l}$  portions). Prior to crystallization, a protein-

DNA-ddCTP sample was prepared at room temperature; approximately 1.2 mg of the 11-nucleotide (nt) template and 0.8 mg of the 6-nt primer (27) were dissolved in 240  $\mu\text{l}$  of a buffer solution (20 mM  $\text{MgCl}_2$ , 0.1 M MES, pH 6.5) and the mixture was left in a sealed Eppendorf tube for 1 hour to allow annealing of the template-primer (28). Two portions of pol  $\beta$  (240- $\mu\text{l}$  of a solution containing 20 mg/ml) were then thawed and added, and the protein-template-primer sample was allowed to stand for an additional hour. A 4- $\mu\text{mol}$  sample of ddCTP (in 40  $\mu\text{l}$  of  $\text{H}_2\text{O}$ ) (29) was the last component to be added, resulting in a reaction mixture containing pol  $\beta$  at approximately 10 mg/ml, 10 mM  $\text{MgCl}_2$ , and an excess of template:primer:ddCTP in molar ratios of 3:4:30, respectively, relative to the amount of protein. The reaction, nucleotidyl transfer of ddCMP to the primer 3' terminus, was allowed to proceed for 2 hours before crystallization trays were set up (30).

Two different crystal forms were observed, depending on the concentration of lithium sulfate in the reservoir solution. One crystal form, obtained with lithium sulfate concentrations from 40 to 75 mM, was hexagonal and grew to dimensions of 0.8 by 0.8 by 0.6 mm in about 2 weeks. These crystals belong to space group  $P6_1$  ( $a = b = 94.9$ ,  $c = 117.6$  Å), with one complex molecule per asymmetric unit. Under similar conditions, but at lithium sulfate concentrations from 75 to 150 mM, platelike crystals grew to dimensions of



**Fig. 2.** Two normal cellular nucleotides, (A) 2'-deoxycytidine 5'-triphosphate (dCTP) and (B) 2'-deoxythymidine 5'-triphosphate (dTTP), and their anti-HIV (drug) counterparts (C) 2',3'-dideoxycytidine 5'-triphosphate (ddCTP) and (D) 3'-azido-2',3'-dideoxythymidine 5'-triphosphate (AZT-TP). The triphosphate moiety, which is linked via a phosphoester bond to the 5' carbon of the ribose, is designated TP.

1.0 by 0.6 by 0.2 mm in a few days. These crystals belong to space group  $P2_1$  ( $a = 106.3$ ,  $b = 56.8$ ,  $c = 86.7$  Å, and  $\beta = 106.4^\circ$ ) and there are two complex molecules per asymmetric unit. Often both crystal forms grew in the same drop, and the crystals on which data were collected (Table 1) were grown at the same concentration of lithium sulfate, 75 mM. The unusually large excess of ddCTP (1:30 molar ratio) was required in order to obtain the  $P6_1$  crystals, but the  $P2_1$  crystals could be grown under much lower ddCTP excesses (1:10 molar ratio). Extreme purity of all components in the crystallization medium, particularly the DNA samples (27), seemed to be an absolute requirement for growing both types of ternary complex crystals.

Attempts were made to obtain ternary complex crystals of rat pol  $\beta$ , a DNA template-primer, and AZT-TP (Fig. 2D) (31) under similar conditions, even though incorporation of AZT-MP would result in a mismatched base pair (of a G-T type) at the primer terminus (22). Orthorhombic crystals grew in space group  $P2_12_12$  ( $a = 188.4$ ,  $b = 67.7$ ,  $c = 39.1$  Å) with one pol  $\beta$  molecule in the asymmetric unit. A 4.0 Å data set was collected and preliminary structural studies (32) showed that, because of crystal packing, it was not possible for the template-primer to occupy the pol  $\beta$  binding channel. Failure of pol  $\beta$  to form a tight complex with the DNA template-primer under these conditions might be attributed to steric hindrance by the azido group of a newly incorporated AZT-MP on the primer terminus.

Efforts to obtain a binary complex of pol  $\beta$  and a DNA template-primer alone (neither ddCTP nor AZT-TP) resulted in crystals that grew under much different condi-

tions, but were nonetheless isomorphous with the  $P2_12_12$  (AZT-TP) crystals mentioned above. Failure of pol  $\beta$  to bind to the DNA in this case could be due to the higher salt concentration of the crystallization medium (about 250 mM salt compared to 75 mM). Because one crystallization medium contained AZT-TP and the other did not, we calculated  $F_{0(AZT-TP)} - F_{0(apo)}$ ,  $\alpha_c$ , difference Fourier maps to see whether an AZT-TP binding site could be located. Strong electron density was observed in an area of the map adjacent to Arg<sup>149</sup>, which is near the catalytically important residues, Asp<sup>190</sup>, Asp<sup>192</sup>, and Asp<sup>256</sup>. This pol  $\beta$ -AZT-TP binary complex which, as discussed below, is probably not catalytically relevant, is similar to a KF-dNTP binary complex in which the dNTP bound to Arg residues near the catalytically important carboxylic acid residues of KF (33).

Human pol  $\beta$ , which has been cloned and expressed (34, 35) in a manner similar to that of rat pol  $\beta$ , shares more than 95 percent sequence similarity with rat pol  $\beta$ , so it was somewhat surprising when attempts to obtain ternary complex crystals of human pol  $\beta$ , a DNA template-primer, and ddCTP under the same conditions described above for the rat enzyme resulted in crystals that grew in two previously unobserved orthorhombic crystal forms. One form has unit cell parameters  $a = 158$ ,  $b = 108$ ,  $c = 60$  Å, with probably two complex molecules in the asymmetric unit, but the crystals diffract only to about 5 Å resolution. In contrast, the second crystal form diffracts quite well (to about 3.3 Å), but its unit cell parameters of  $a = 465$ ,  $b = 168$ ,  $c = 56$  Å are so large that special data collection techniques would be required.

**Structure determination and refinement.** Data collection and refinement statistics for the structure determinations of the two ternary complexes of rat pol  $\beta$ , a DNA template-primer, and ddCTP are listed in Table 1. Structure solutions utilized the refined atomic coordinates of the high resolution (2.3 Å) structure of the 31-kD domain of rat pol  $\beta$  (14). The molecular replacement programs of XPLOR (36) gave clear rotation solutions for the 31-kD domain of both ternary complexes, but only after results from classical cross-rotation searches had been filtered through the Patterson-correlation (PC) refinement steps (37). PC refinement techniques were particularly powerful for our structure determinations because independent rigid body movements of the fingers, palm, and thumb subdomains of the 31-kD domain could be allowed during PC refinement of the rotation searches. Results from subsequent translation searches gave solutions for the  $P6_1$  complex structure that were, in general, at higher peak height to background ratios than translation solutions for the  $P2_1$  complex structure, but the highest translation peaks in both cases nevertheless were the correct solutions (38).

The 31-kD partial structure solutions obtained by molecular replacement techniques were subjected to rigid body refinement by XPLOR (36), where the entire 31-kD domain was first allowed to move as a rigid body, then later, the fingers, palm, and thumb subdomains were allowed to move as independent rigid bodies simultaneously. Typical R factors at this stage were about 50 percent. After subsequent positional refinement with the least squares program package TNT (39) had lowered the R factors of the partial solutions to about 45 percent, we calculated  $F_0 - F_c$ ,  $\alpha_c$ , difference Fourier maps that revealed clear electron density for many of the backbone phosphates of a double-stranded DNA molecule as well as the three phosphates of a ddCTP nucleotide, and even portions of the 8-kD domain were evident at this early stage. Cycles of model building and least squares refinement improved the electron density for the rest of the DNA as well as the 8-kD domain for both complex structures, and once the R factors had dropped below 30 percent, refinement of individual isotropic temperature factors also improved the maps and facilitated refinement.

Although we were unable to discern the DNA base sequences at these resolutions, the directionality (5'  $\rightarrow$  3') of the DNA strands was evident early in our modeling efforts, hence we knew that the 3' terminus of either the template strand or the primer strand was positioned at the pol  $\beta$  active site. What ultimately distinguished the template from the primer was that we were

**Table 1.** Data collection and refinement statistics. X-ray diffraction data were collected on a multiwire area detector (98) (San Diego Multiwire Systems) with monochromatized  $\text{CuK}\alpha$  radiation (Rigaku rotating anode x-ray generator), and intensity observations for each data set were processed with a local UCSD Data Collection Facility software package (99). Reflections from 20 Å to the maximum resolution were included in all least squares refinement steps. The final structures for both complexes include all residues, with the exception of residues 1 to 8 of the disordered  $\text{NH}_2$ -terminus and residues 246 to 248 of a disordered surface loop. There are a few missing side chain atoms in both coordinate sets that are mainly in lysine and arginine residues of the 8-kD domain. Omit maps were used to confirm the modeling of the DNA template-primer, the ddCTP nucleotide, and the cis-peptide bond between Gly<sup>274</sup> and Ser<sup>275</sup>.

Space group	$d_{\min}$ (Å)	$\frac{1}{\sigma^*}$	Data collection		Completeness (%)	$R_{\text{sym}}^\dagger$	Atoms $^\ddagger$	Refinement		Final $R_{\parallel}$
			Reflections					rms deviation $^\S$		
			Total	Unique				Bond (Å)	Angle ( $^\circ$ )	
$P6_1$	2.9	1.8	53,583	13,281	99	0.087	2914	0.020	2.9	0.193
$P2_1$	3.6	1.8	25,046	10,650	96	0.059	5753	0.018	2.9	0.199

\*Average ratio of observed intensity to background in the highest resolution shell of reflections.  $^\dagger R_{\text{sym}} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{avg}}$ .  $^\ddagger$ The number of nonhydrogen atoms includes 31 and 4 water oxygens for the  $P6_1$  and the  $P2_1$  structures, respectively.  $^\S$ The rms bond and rms angle values are the deviations from ideal values of the bond lengths and bond angles in the final model.  $^\parallel$ Final  $R = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$ , including all data between 20 Å and the maximum resolution.

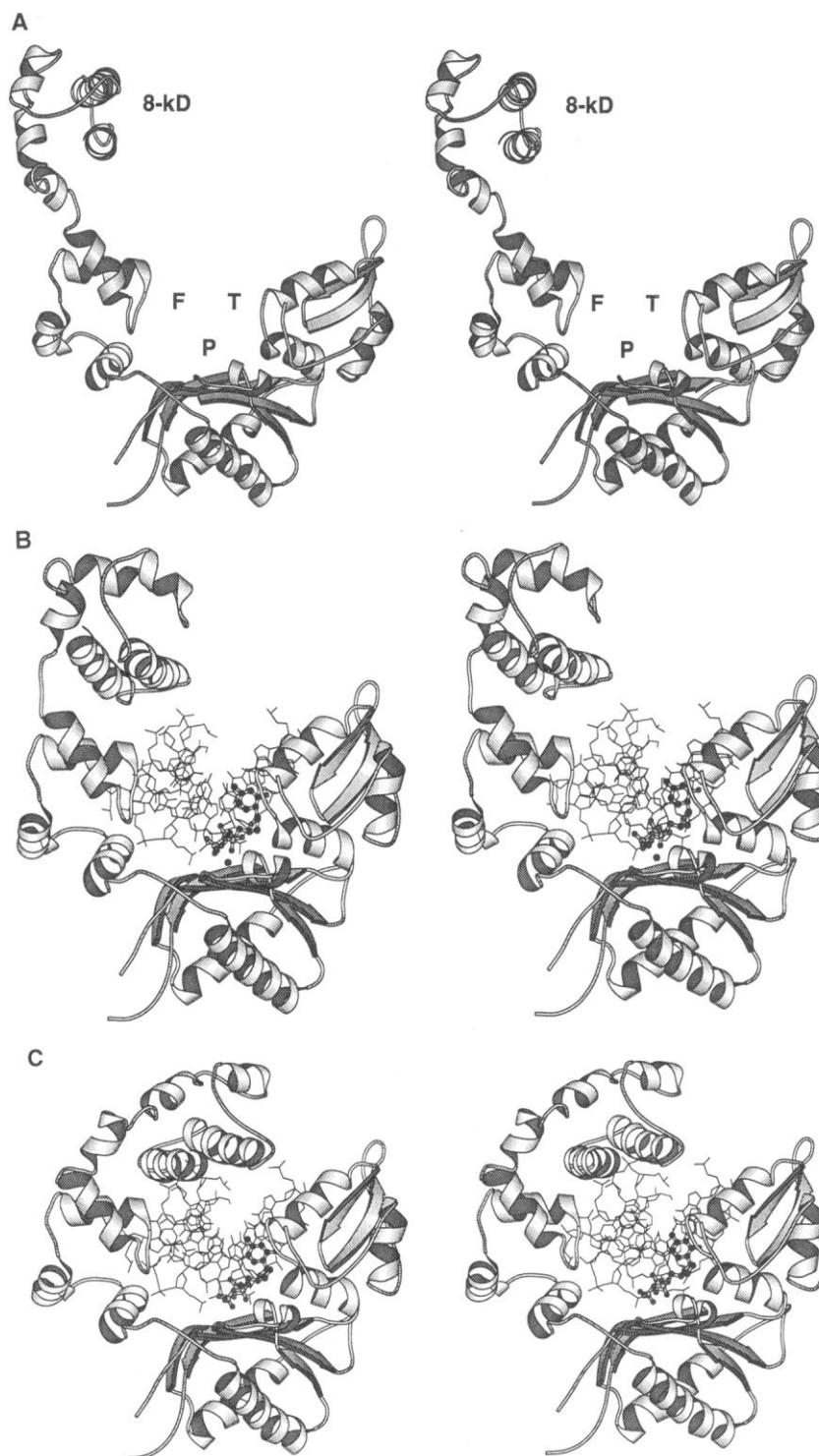
able to model in seven nucleotides for one of the DNA strands and at least 8 nt for the second DNA strand. Provided that no unexpected side reactions had occurred during crystallization, we knew that the primer could be no longer than 7 nt, and therefore the DNA strand containing 8 nt was designated the template. We then concluded that the first three bases of the template (AAA) are disordered in both crystal structures. This interpretation of the data is in agreement with the idea that the 3' terminus of the primer should be positioned at the polymerase active site. Analysis of the DNA in our refined structures with the program CURVES II (40) indicated that the DNA is predominantly B form. Our DNA may have some A-DNA characteristics, however, in that the minor groove width appears to increase as the DNA approaches the pol  $\beta$  active site; the section of the double-stranded DNA that is removed from the active site and protrudes into solution is characteristic of B-DNA with a minor groove width of 11 Å, whereas nearer to the active site, the minor groove width is almost 15 Å (typical A-DNA has a minor groove width of about 17 Å).

**Description of the structures.** When the pol  $\beta$  ternary complex structures are compared with the structure of the apo enzyme (Fig. 3), the most apparent differences consist of large movements of the 8-kD NH<sub>2</sub>-terminal domain relative to the fingers, palm, and thumb of 31-kD COOH-terminal domain. The 8-kD domain is tethered to a proteolytically sensitive hinge region (residues 80–90) and changes from an open conformation in the apo structure to more closed conformations in the complex structures. Because of the precarious position of the 8-kD domain in the pol  $\beta$  apo structure, this type of conformational change seemed inevitable even before the structures of the ternary complexes were determined. The only other significant conformational changes on complex formation were noticeable rigid-body movements of the thumb and, to a lesser degree, the fingers, resulting in a somewhat more tightly closed hand in the ternary complexes. A greater degree of flexibility on the part of the thumb subdomain has also been observed in other polymerase-DNA structures (12, 23, 24, 41). A least squares superposition of the 31-kD domain of the apo structure (14) on the 31-kD domain of one of the ternary complex structures (*P*<sub>61</sub>) resulted in a root-mean-square (rms) deviation in  $\alpha$  carbon positions of 2.5 Å, whereas when the fingers, palm, and thumb subdomains were treated separately, the rms deviations in  $\alpha$  carbon positions were only 0.71, 0.69, and 0.82 Å, respectively.

The 8-kD domain has a net charge of +10 (assuming neutral histidines) and

binds to single stranded DNA with an association constant of  $2 \times 10^5 \text{ M}^{-1}$  (26). It has no obvious structural equivalent in any of the other polymerases for which

crystal structures have been determined, and crosslinking studies with gapped DNA substrates (42) suggest that the 8-kD domain is probably responsible for the highly



**Fig. 3.** Stereoview ribbon diagrams (100) of (A) rat DNA pol  $\beta$ , apo structure (14) and (B and C) ternary complexes of rat DNA pol  $\beta$  with a DNA template-primer and ddCTP in space groups *P*<sub>61</sub> and *P*<sub>21</sub>, respectively. In (A) the 8-kD domain is designated 8-kD, and the fingers, palm, and thumb subdomains of the 31-kD domain are represented by F, P, and T, respectively. A ball-and-stick representation highlights the ddCTP nucleotide in (B and C). In (B) the positions of the two Mg<sup>2+</sup> ions are marked with black spheres. These metals ions are not shown in (C) because we were unable to see the Mg<sup>2+</sup> ions in electron density maps of the lower resolution *P*<sub>21</sub> ternary complex structure.

processive short-gap filling activities found exclusively in pol  $\beta$  (43). It has been proposed that the 31-kD domain binds to the double-stranded segment of the template-primer, and the 8-kD domain binds to the single-stranded template overhang (44)—or in the case of binding to a short gap in the DNA, to the 5'-phosphate of the downstream oligonucleotide (42). We see some evidence of this in that the 31-kD domain clearly uses its palm, fingers, and thumb to grasp the double-stranded segment of the template-primer while the 8-kD domain, although positioned quite differently in the two complex structures, is nevertheless close to where an extended template would be. Unfortunately, our tem-

plate overhang was probably a little too short (only four bases—GAAA) to interact strongly with the 8-kD domain, causing the first three bases of the template to be disordered in both crystal structures. It is possible that, because the highly flexible 8-kD domain had no template on which to anchor in our crystallization experiments, its position was determined almost entirely by crystal packing forces, and probably neither of the two conformations of the 8-kD domain seen in Fig. 3, B and C, is correct for template binding *in vivo*. Nevertheless, kinetic studies of the 31-kD fragment alone showed that pol  $\beta$  can still function as a polymerase without the 8-kD domain, albeit at only about 5 percent of its normal activity (44).

In contrast to the 8-kD domain, the rest of the structure (the fingers, palm, and thumb of 31-kD domain, as well as the template-primer and ddCTP substrate) is virtually identical in both crystal forms of the ternary complex. This provides support for the physiological relevance of our complex crystals, at least with respect to the polymerase-DNA-ddCTP interactions. Also strengthening the argument is that, unlike other reported crystals of polymerase-DNA complexes (12, 41), our crystals were grown at low, near physiological salt concentrations. Finally, the fact that a ddCMP was incorporated into our template-primer shows that the nucleotidyl transfer reaction did proceed, at least for one turnover, in the same medium from which crystals were eventually obtained. In that the following discussions do not apply to the 8-kD domain of pol  $\beta$  and will be limited mostly to the 31-kD domain's interactions with DNA and ddCTP, we will henceforth refer only to the complex structure that has been refined to the highest resolution, the  $P6_1$  crystal structure.

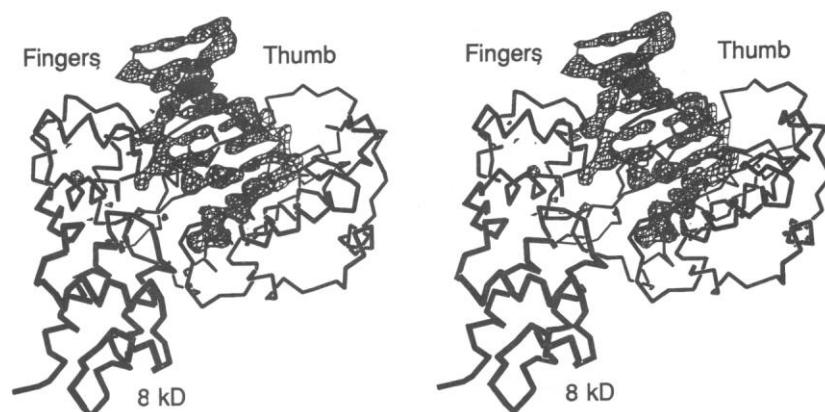
The DNA binding channel in pol  $\beta$ , just as in KF, RT, and RNAP, is lined with positively charged lysine and arginine side chains, and it has always been a reasonable assumption that their function is to stabilize the negatively charged backbone phosphates of the DNA (45). Therefore it was quite surprising that, except for Arg<sup>254</sup>, which is hydrogen bonded to the phosphate of the newly incorporated ddCMP of the primer strand, there are no direct lysine or arginine interactions with the backbone phosphates of the DNA in our complex (Table 2). Instead, nearly all of the interactions of protein with DNA involve two different clusters of protein backbone nitrogens located at the entrance to the DNA binding channel (Table 2). One cluster, consisting of four of the backbone nitrogens between Gly<sup>105</sup> and Ala<sup>110</sup>, is located at the NH<sub>2</sub>-terminal end of helix G in the fingers subdomain of pol  $\beta$  and interacts with the phosphates of the primer strand. The second cluster, comprising three of the five backbone nitrogens between Lys<sup>230</sup> and Lys<sup>234</sup>, is located in a beta turn, which connects beta strands 3 and 4 of the palm subdomain and interacts with backbone phosphates of the template strand. The only other hydrogen bonded interactions (3.3 Å or less) between pol  $\beta$  and DNA phosphates are between the side chains of Thr<sup>292</sup> and Tyr<sup>296</sup>, located on a loop between beta strands 6 and 7 of the thumb subdomain, and the backbone phosphates of the template strand (Table 2).

In addition to our observations that there seemed to be fewer hydrogen bond interactions between pol  $\beta$  and DNA than expected (Table 2), we were also initially

**Table 2.** Hydrogen bond interactions of 3.3 Å or less between pol  $\beta$  and the DNA template-primer.

Residue	Subdomain	Atom	Base*	Atom	Distance (Å)
<i>Protein to DNA phosphate H bonds</i>					
Gly <sup>105</sup>	Fingers	N	P-6C	O2P	2.9
Gly <sup>107</sup>	Fingers	N	P-5G	O2P	2.7
Ser <sup>109</sup>	Fingers	N	P-5G	O1P	2.9
Ala <sup>110</sup>	Fingers	N	P-5G	O2P	3.1
Arg <sup>254</sup>	Fingers	NH2	P-7C	O2P	2.7
Lys <sup>230</sup>	Palm	N	T-9C	O2P	3.0
Thr <sup>233</sup>	Palm	N	T-8G	O2P	3.1
Lys <sup>234</sup>	Palm	N	T-8G	O2P	2.7
Thr <sup>292</sup>	Thumb	OG1	T-5G	O2P	2.7
Tyr <sup>296</sup>	Thumb	OH	T-5C	O2P	2.6
<i>Protein to DNA base H bonds</i>					
Lys <sup>234</sup>	Palm	NZ	T-7C	O2	2.9
Tyr <sup>271</sup>	Thumb	OH	P-7C	O2	2.7
Arg <sup>283</sup>	Thumb	NH1	T-4G	N3	3.2

\*DNA bases are designated T or P to distinguish template bases from primer bases, respectively. Starting from the 5' terminus of each strand, bases are numbered 1 to 11 for the template and are numbered 1 to 7 for the primer. C and G represent cytosine and guanine, respectively, and atom designations follow Protein Data Bank nomenclature.



**Fig. 4.** Omit map of the DNA template-primer and ddCTP overlaid on an  $\alpha$  carbon diagram of the refined  $P6_1$  pol  $\beta$  structure. The view is that of Fig. 3 rotated by 90° about a horizontal axis in the plane of the page so that the DNA binding channel is now vertical. The template-primer sits on top of the palm subdomain, which is not labeled. Before all omit maps were calculated, the part of the structure in question was deleted from the coordinate file and the remaining partial structure was subjected to 200 cycles of least squares positional refinement in XPLOR (36) in order to remove bias from the phases.

surprised to see that the DNA sits in the binding channel at a slight angle and appears to "run into" alpha helices M and N of the thumb subdomain (Fig. 4). It is possible that the angle between the DNA axis and the apparent axis of the pol  $\beta$  binding channel would change considerably if a longer template were used and, as proposed above, the 8-kD domain participated in the positioning of the template-primer. However, the aesthetically pleasing observation that the base pairs of the DNA are parallel to the beta strands of the palm subdomain (Fig. 5) encourages us to believe that interactions of pol  $\beta$  with the double-stranded segment of any DNA template-primer will not vary much from what is seen in the present ternary complex structures.

Perhaps one of the most unvarying characteristics of B-DNA is that it has a spine of well-ordered water molecules which interacts with the O2 of pyrimidines and the N3's of purines in the minor groove, and it has been proposed that the disruption of this particular water structure is the first step in the B-DNA to A-DNA transition (46, 47). In our complex structure, only three protein side chains come within 3.3 Å of the DNA bases, and they are all located in the shallow minor groove of the template-primer (Table 2). Two of these (Lys<sup>234</sup> and Tyr<sup>271</sup>) are hydrogen bonded to the O2 of a template cytidine and the O2 of a primer cytidine, respectively, while another (Arg<sup>283</sup>) is hydrogen bonded to the N3 of a template guanine. This leads us to propose that perhaps Lys<sup>234</sup>, Tyr<sup>271</sup>, and Arg<sup>283</sup> all function to break up the water structure in the minor groove of the template-primer upon complex formation, resulting in a larger minor groove width, characteristic of A-DNA, at the pol  $\beta$  active site.

Unlike transcription factors and other gene regulatory DNA-binding proteins, polymerases must bind to DNA with little regard for sequence specificity. This is evident from Table 2 where, as discussed above, most of the protein to DNA interactions are nonsequence-specific hydrogen bonds between pol  $\beta$  backbone nitrogens and DNA phosphate oxygens. Even what appear to be sequence-specific interactions between protein side chains and DNA bases, upon closer inspection turn out to be rather nonspecific in that each of the three protein side chains mentioned above—Lys<sup>234</sup>, Tyr<sup>271</sup>, and Arg<sup>283</sup>—that are in contact with DNA bases can act as an unbiased hydrogen bond donor to either the O2 of pyrimidines or the N3 of purines in the minor groove.

**Description of the active site.** Much of the ddCTP binding pocket in the ternary complex is made up of the 3' terminus of the primer strand (Fig. 6A, toward the left)

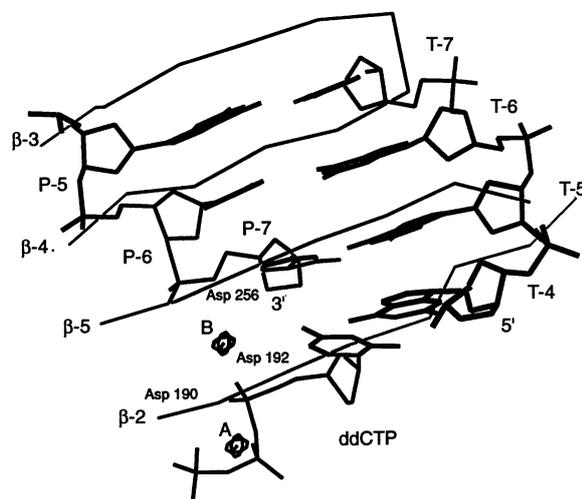
and the template overhang (at the top), where the base pairing of ddCTP with a complementary base, guanine, is evident in the crystal structure (Fig. 6A and Table 3). The idea that the template-primer makes up a large part of the binding site in pol  $\beta$  for the incoming ddCTP is consistent with kinetic studies showing that there is a strict kinetic order of binding for the substrates in nucleotidyl transfer reactions catalyzed by pol  $\beta$  (48), *E. coli* pol I (49), and RT (50), the polymerase binds to the template-primer first, and then nucleotide binding takes place. Supporting this view, structural studies of a pol  $\beta$ -dATP binary complex (14) and of a KF-dATP binary complex (33), both of which were crystallized in the absence of a template-primer, revealed nucleotide binding sites that differed somewhat from that seen in our ternary complex.

Making up most of the right side of the nucleotide binding pocket and interacting primarily with the base and ribose moieties of the ddCTP is a structural motif consisting of a sharp kink, made possible by a cis-peptide bond, between  $\alpha$  helices M and N in the thumb subdomain of pol  $\beta$  (Fig. 6A). This was the only cis-peptide found in the pol  $\beta$  apo structure (between Gly<sup>274</sup> and Ser<sup>275</sup>) (14) and even though it appears to have remained as a cis-peptide upon complex formation with DNA, its proximity to the active site nevertheless suggests that it may play a dynamic role in catalysis. For instance, it is evident that a cis- to trans-peptide bond transition, perhaps occurring during or just after the incorporation of a nucleotide, could result in a large displacement of one or both helices (M and N), which in turn could push the enzyme off the template-primer. Thus, we propose that the cis-peptide in pol  $\beta$  may function to facilitate the product-off step of catalysis, which is typically the steady-state rate-limiting step for polymerases when in a distributive mode of synthesis (21, 51, 52). Another

possibility is that the cis-peptide bond, which links the active site to the template overhang via helix N (Fig. 6A), functions only to facilitate the translation of the enzyme along the template during processive polymerization.

It is quite appropriate that the protein side chain that interacts specifically with the base of ddCTP, Asn<sup>279</sup> of helix N (Table 3), is unbiased toward all four possible incoming nucleotides in that it can act either as a hydrogen bond acceptor or a hydrogen bond donor. The only other interactions (of 4 Å or less) between pol  $\beta$  and the base moiety of ddCTP are nonspecific van der Waals contacts between the side chain carbon (CB) of Asp<sup>276</sup> and ring carbon atoms (C4 and C5) of the cytidine base (Table 3).

Pol  $\beta$  interacts intimately with the sugar moiety of ddCTP as is evident from the close van der Waals contacts between the protein backbone atoms of Tyr<sup>271</sup>, Phe<sup>272</sup>, and Gly<sup>274</sup> (53), and the ribose ring carbons, C2' and C3', of ddCTP (Fig. 6A and Table 3). Because the only difference between a ribonucleotide and a deoxyribonucleotide is a hydroxyl at the C2' of the ribose ring, the protein backbone segment Tyr<sup>271</sup>-Gly<sup>274</sup> may participate in nucleotide selectivity of DNA over RNA for pol  $\beta$ . For example, a DNA polymerase can be converted to a relatively efficient RNA polymerase by changing the metal ion in the reaction medium from Mg<sup>2+</sup> to Mn<sup>2+</sup> (54). The two metal ions in the pol  $\beta$  active site together bind to all three phosphates ( $\alpha$ ,  $\beta$ , and  $\gamma$ ) of ddCTP (Table 3), and therefore the metal ions probably participate to some extent in the positioning of the incoming nucleotide. As seen in our structure, only a slight change in ddCTP orientation, which could be induced by a change from Mg<sup>2+</sup> to Mn<sup>2+</sup> ions in the active site, would be required to reduce steric hindrance at the C2' ribose of ddCTP, possibly making



**Fig. 5.** Alignment of the template-primer base pairs with the beta strands of the palm subdomain. The view is the same as that in Fig. 4. Beta strands are labeled  $\beta$ -2 through  $\beta$ -5. The  $\alpha$  carbon positions of the catalytically important residues, Asp<sup>190</sup>, Asp<sup>192</sup>, and Asp<sup>256</sup>, are shown, as well as the nucleotide substrate, ddCTP. The two Mg<sup>2+</sup> ions in metal sites A and B are labeled A and B, respectively. Base designations defined in Table 2.

RNA just as good a substrate as DNA.

Another example of nucleotide selectivity, this time dTTP over AZT-TP, might also be explained by steric hindrance in the pol  $\beta$  active site, particularly at the ribose C3' (Fig. 2B), binds more tightly to a pol  $\beta$ -DNA complex than does its analog AZT-TP, which has a bulky azido group at C3' (Fig. 2D) (55). Such observations might explain why a drug like AZT specifically targets the HIV-1 RT and not host cell polymerases like pol  $\beta$ ; in contrast to pol  $\beta$ , RT shows no selectivity in binding dTTP compared to AZT-TP during reverse transcription (56), and thus, perhaps RT lacks the structural equivalent of the pol  $\beta$  "selective" Tyr<sup>271</sup> to Gly<sup>274</sup> backbone segment, making RT more susceptible to AZT-TP inhibition.

Three of the six hydrogen bonds between the protein and the negatively charged phosphate moiety of ddCTP involve nitrogen backbone atoms of pol  $\beta$  (Table 3). Although this is reminiscent of

the nitrogen clusters mentioned above that help to stabilize the template and primer phosphates in the DNA binding channel, the geometry of the nitrogen backbones differ in that they more closely resemble the mononucleotide binding motifs found in other enzymes. This and other aspects of the interactions between protein and ddCTP phosphates in the pol  $\beta$  active site are discussed in (14).

The strongest interactions with ddCTP in the pol  $\beta$  active site, however, are not with protein side chains directly, but rather with two Mg<sup>2+</sup> ions that in turn coordinate the side chain oxygens of Asp<sup>190</sup>, Asp<sup>192</sup>, and Asp<sup>256</sup> (Fig. 6 and Table 3). These three carboxylic acids are present in all DNA polymerases for which amino acid sequences are known (15), and mutagenesis studies show that they are critical for catalysis in pol  $\beta$  (57), KF (58, 59), and RT (60). However, we avoid referring to these three aspartic acids as a "catalytic triad" because only two of the three (Asp<sup>190</sup> and Asp<sup>256</sup>) are present in all known RNA polymerase amino acid sequences (15). As

discussed above, converting a DNA polymerase to a relatively efficient RNA polymerase can be as simple as changing the metal ions in the active site from Mg<sup>2+</sup> to Mn<sup>2+</sup> (54), and consequently there is not much reason to suppose that the RNA polymerases utilize a nucleotidyl transfer mechanism that differs drastically from the mechanism proposed below for a DNA polymerase such as pol  $\beta$ . In support of this view, mutagenesis experiments that targeted catalytically important residues in KF (59) showed that Asp<sup>882</sup> and Asp<sup>705</sup> (the equivalents of Asp<sup>190</sup> and Asp<sup>256</sup> in pol  $\beta$ ) are much more critical to catalysis than the third carboxylic acid of the trio, Glu<sup>883</sup> (equivalent to Asp<sup>192</sup> in pol  $\beta$ ), which is not conserved in the RNA polymerases. In agreement with these observations and as discussed below, the primary function of Asp<sup>256</sup> in the pol  $\beta$  active site appears to aid in stabilizing the pentacoordinated  $\alpha$  phosphate of the transition state, whereas the primary functions of Asp<sup>190</sup> and Asp<sup>192</sup> are to aid in positioning the nucleotide substrate. This suggests that pol  $\beta$  could quite possibly still function in the absence of one of the two latter carboxylic acids, such as Asp<sup>192</sup>, which would be the case for an RNA polymerase. Perhaps also significant is the observation that the metal-to-oxygen bonds are longer for Asp<sup>192</sup> than they are for Asp<sup>190</sup> in the pol  $\beta$  active site (Table 3).

**Mechanism of nucleotidyl transfer.** In 1979, with results from kinetic experiments on *E. coli* pol I that utilized phosphorothioate dATP analogs, Burgers and Eckstein proposed that the pol I catalyzed nucleotidyl transfer reaction had the following properties (61): (i) A divalent metal ion (Mg<sup>2+</sup>) is bound specifically to the  $\beta$  and  $\gamma$  phosphates of the nucleotide (dATP) as a  $\beta, \gamma$ -bidentate; (ii) the negative charge on the  $\alpha$ -phosphate of the nucleotide is neutralized by a positive group on the enzyme; and (iii) attack by the 3'-OH group of the primer on the  $\alpha$  phosphate and subsequent release of the PP<sub>i</sub> (or Mg-PP<sub>i</sub>) proceeds in an in-line fashion.

All three of these properties are in good agreement with the geometry seen in the active site of our ternary complex, and in fact, our active site (Fig. 6) does not differ significantly from that proposed by Burgers and Eckstein [figure 4 of (61)]. In agreement with criterion number (i), the pol  $\beta$  active site shows a Mg<sup>2+</sup> ion (metal site A) bound as a bidentate to the  $\beta$  and  $\gamma$  phosphates of ddCTP (Table 3). As for criterion number (ii), we now know that the originally undefined "positive group on the enzyme" that stabilizes the  $\alpha$  phosphate of the nucleotide is simply the second Mg<sup>2+</sup> ion in metal site B (Fig. 6). In fact, the geometry of the Mg<sup>2+</sup> ion in site B of our

**Table 3.** Interatomic distances of interest for the pol  $\beta$  active site.

ddCTP moiety	Atom	Base* or residue	Atom	Distance (Å)
<i>Contacts between ddCTP and DNA or protein</i>				
Base	N4	T-4 G	O6	3.0
	N3	T-4 G	N1	2.7
	O2	T-4 G	N2	2.7
	C4	Asp <sup>276</sup>	CB	3.5
	C5	Asp <sup>276</sup>	CB	3.7
	O2	Asn <sup>279</sup>	ND2	3.0
Ribose	C2'	Tyr <sup>271</sup>	O	3.5
	C3'	Phe <sup>272</sup>	O	3.2
	C2'	Gly <sup>274</sup>	CA	3.2
$\alpha$ Phosphate	PA	P-7 C	O3'	4.3†
$\beta$ Phosphate	O1B	Ser <sup>180</sup>	N	3.0
	O1B	Arg <sup>183</sup>	NH2	2.7
	O3B	Ser <sup>180</sup>	OG	2.8
$\gamma$ Phosphate	O1G	Gly <sup>189</sup>	N	2.8
	O1G	Asp <sup>190</sup>	N	3.0
	O2G	Arg <sup>149</sup>	NH2	2.5
	Ligand	Atom	Distance (Å)	
Mg <sup>2+</sup> site A	Asp <sup>190</sup>	OD1	2.0	
	Asp <sup>192</sup>	OD2	2.7	
	ddCTP, $\beta$ phosphate	OD2	1.7	
	ddCTP, $\gamma$ phosphate	O1G	2.3	
	Water	O	2.6	
Mg <sup>2+</sup> site B	Asp <sup>190</sup>	OD2	2.0	
	Asp <sup>192</sup>	OD1	2.6	
	Asp <sup>256</sup>	OD1	3.0	
	ddCTP, $\alpha$ phosphate	O1A	2.7	
	P-7 C	O3'	2.9†	
	Water	O	—‡	

\*Base designations are in Table 2. †These distances were obtained with a O3' atom that had to be modeled into the active site because the newly incorporated ddCMP primer terminus lacks this group. ‡This water is not seen in our crystal structures, but because the Mg<sup>2+</sup> ion in site B has octahedral geometry, we propose that a water might occupy this empty sixth-ligand position.

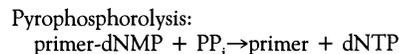
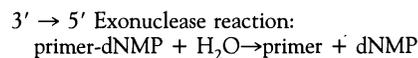
structure can explain why polymerases are highly selective for only the S, as opposed to the R, absolute configuration of dATP $\alpha$ S phosphorothioate analogs whenever Mg<sup>2+</sup> is the metal ion in the reaction mixture (61). Mg<sup>2+</sup> coordinates oxygen much more strongly than sulfur, and in our active site the Mg<sup>2+</sup> ion in metal site B coordinates a specific, nonesterified oxygen of the  $\alpha$ -phosphate of the nucleotide. If this particular oxygen were replaced by a sulfur, as would be the case for the R configuration of dATP $\alpha$ S, coordination by the Mg<sup>2+</sup> ion in site B should be weakened considerably, at least according to the geometry our structure. Finally, the geometry of the active site is also in accord with an in-line mechanism, criterion number (iii), for the nucleotidyl transfer reaction. An in-line mechanism, proposed because the polymerase reaction proceeds with inversion of configuration at the  $\alpha$  phosphate (61), restricts the possible orientation of the attacking group with respect to the leaving group. The

attacking and leaving groups must be opposite one another, relative to the  $\alpha$  phosphate, in order to occupy the two apical positions of the pentacoordinated  $\alpha$  phosphate in the transition state. In the pol  $\beta$  active site, the 3' carbon of the primer strand (which normally possesses the attacking 3'-OH) lies just opposite, relative to the  $\alpha$  phosphate, to the scissile oxygen of the PP<sub>i</sub> leaving group.

It is not too surprising that the active site of pol  $\beta$  is similar to the active site of the 3'  $\rightarrow$  5' exonuclease domain of *E. coli* pol I, which is known, through extensive structural (23, 24) and mutagenesis (62, 63) studies, to employ a two-metal ion mechanism for phosphoryl transfer, and like pol  $\beta$ , proceeds with inversion of configuration at the scissile phosphate (64). Like the 3'  $\rightarrow$  5' exonuclease, pol  $\beta$  has three carboxylic acids that position two divalent metal ions about 4 Å apart in the active site and, although the orientations of the carboxylic acids are quite different in

the two structures, the geometries of the metal sites are strikingly similar. The metal ion in site A in both cases tightly coordinates four oxygen ligands with highly distorted tetragonal geometry and has a fifth, weakly bound water ligand (Table 3). In fact, a better description of the unusual geometry around metal site A in our structure would be that of a square pyramid in which the Mg<sup>2+</sup> is at the apex, rising about 1.8 Å out of the square plane described by four oxygen ligands (one from Asp<sup>190</sup>, one from Asp<sup>192</sup>, one from the  $\beta$  phosphate, and one from the  $\gamma$  phosphate) (Fig. 6) (Table 3). In contrast, the metal ion in site B (for both pol  $\beta$  and the 3'  $\rightarrow$  5' exonuclease) is not as tightly bound as the metal ion in site A and has slightly distorted octahedral geometry (65). Just as with the 3'  $\rightarrow$  5' exonuclease, we propose that the primary function of the metal ion in site A is to aid in binding and positioning of the substrate, and the primary function of the metal ion in site B is to help stabilize the pentacoordinated phosphate of the transition state, though both metal sites probably participate in both of these functions to some degree. Finally, as is the case with the 3'  $\rightarrow$  5' exonuclease, we propose that a metal ion activates the attacking oxygen by acting as a Lewis acid, while a protein side chain acts as a proton acceptor (Fig. 6B). For the 3'  $\rightarrow$  5' exonuclease, the metal ion in site A activates an attacking water molecule, while Glu<sup>357</sup> acts as the proton acceptor and, in the case of pol  $\beta$ , the metal ion in site B activates the attacking 3'-OH of the primer, while Asp<sup>256</sup> acts as the proton acceptor. All of these features of the transition state of the nucleotidyl transfer reaction of pol  $\beta$  are evident from the crystal structure, and only the missing 3'-OH of the primer strand was added in order to draw the schematic in Fig. 6B.

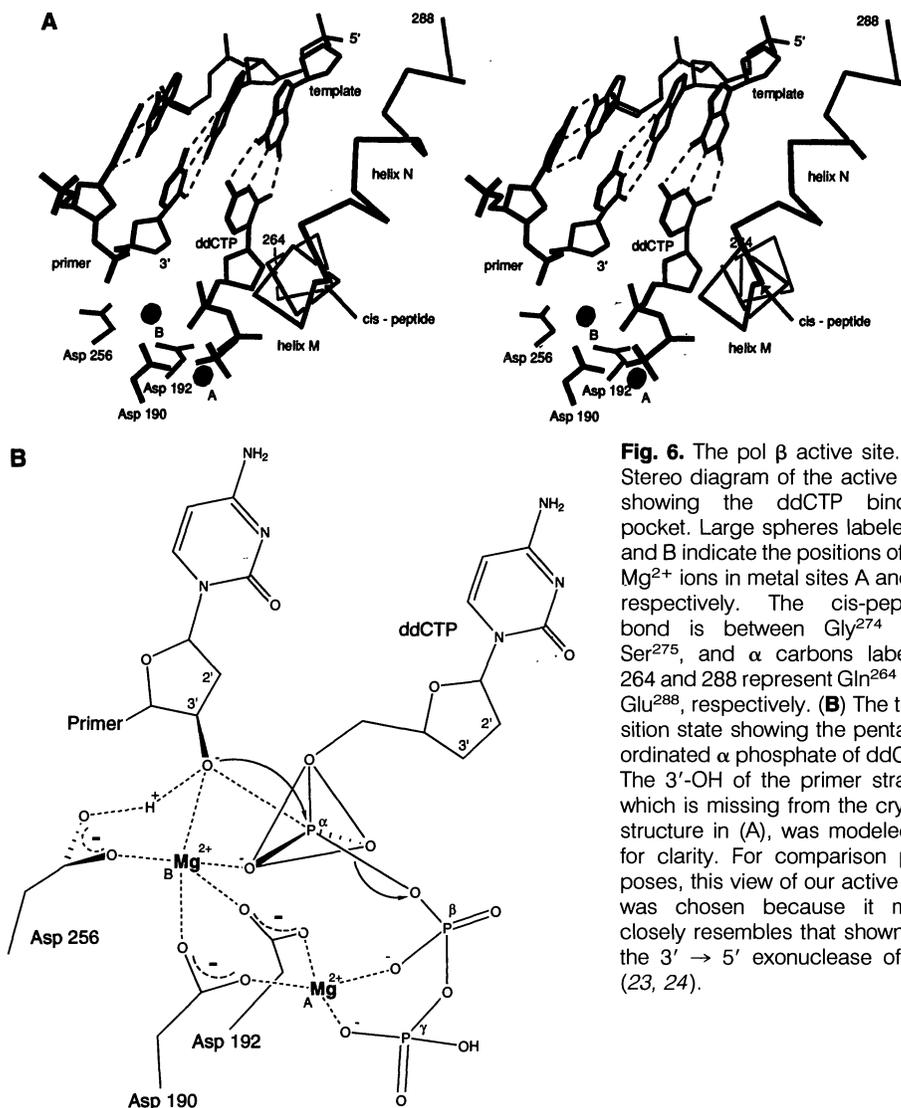
Despite all the similarities, we caution against thinking of the 3'  $\rightarrow$  5' exonuclease reaction as "the polymerase reaction in reverse." In fact, the polymerase reaction in reverse is pyrophosphorolysis and differs from the exonuclease reaction as follows:



so that the metal ion in site A for the 3'  $\rightarrow$  5' exonuclease interacts primarily with a water molecule (24), whereas the metal ion in site A for the polymerase interacts primarily with PP<sub>i</sub> (or the PP<sub>i</sub> moiety of ddCTP) (65).

In summary, the overall nucleotidyl transfer reaction, as catalyzed by pol  $\beta$ , probably proceeds as follows (Fig. 6):

i) As pol  $\beta$  binds to the template-primer, Lys<sup>234</sup>, Tyr<sup>271</sup>, and Arg<sup>283</sup> break up the



**Fig. 6.** The pol  $\beta$  active site. **(A)** Stereo diagram of the active site showing the ddCTP binding pocket. Large spheres labeled A and B indicate the positions of the Mg<sup>2+</sup> ions in metal sites A and B, respectively. The cis-peptide bond is between Gly<sup>274</sup> and Ser<sup>275</sup>, and  $\alpha$  carbons labeled 264 and 288 represent Gln<sup>264</sup> and Glu<sup>288</sup>, respectively. **(B)** The transition state showing the pentacoordinated  $\alpha$  phosphate of ddCTP. The 3'-OH of the primer strand, which is missing from the crystal structure in (A), was modeled in for clarity. For comparison purposes, this view of our active site was chosen because it most closely resembles that shown for the 3'  $\rightarrow$  5' exonuclease of KF (23, 24).

water structure of the minor groove, causing the minor groove width to increase near the active site.

ii) An incoming nucleotide is positioned in the active site by base-pairing with the template, by a hydrogen bond to Asn<sup>279</sup>, by a van der Waals contacts with Asp<sup>276</sup>, by steric hindrance with the protein backbone between Tyr<sup>271</sup> and Gly<sup>274</sup>, by six hydrogen bonds between protein and ddCTP phosphates, and by two Mg<sup>2+</sup> ions that are, in turn, positioned by Asp<sup>190</sup>, Asp<sup>192</sup>, and Asp<sup>256</sup>.

iii) The Mg<sup>2+</sup> ion in metal site B, acting as a Lewis acid, activates the 3'-OH of the primer terminus, while one of its ligands, Asp<sup>256</sup>, probably acts as the proton acceptor for the 3'-OH.

iv) After attack on the  $\alpha$ -phosphate by the activated 3'-OH, the reaction proceeds through the transition state in which the  $\alpha$  phosphate is pentacoordinated, with the attacking 3' oxygen of the primer terminus and the leaving oxygen of the PP<sub>i</sub> group occupying the two apical positions. The pentacoordinated transition state is stabilized by the Mg<sup>2+</sup> ion in site B, which coordinates both an apical and an equatorial oxygen of the  $\alpha$  phosphate.

v) After PP<sub>i</sub> (or Mg-PP<sub>i</sub>) is released, pol  $\beta$  is ready for another cycle, but only after the enzyme either releases from the template-primer (distributive mode of synthesis) or pulls itself along the template (processive mode of synthesis); one or both of these activities may be facilitated by a conformational change of cis- to trans-peptide at Gly<sup>274</sup>-Ser<sup>275</sup>.

Metal ions play crucial roles in the mechanism of pol  $\beta$  and of the 3'  $\rightarrow$  5' exonuclease of pol I. This kind of independence from direct involvement of protein side chains during catalysis has led to proposals that perhaps hydrolysis reactions involving nonprotein catalysts such as self-splicing ribozymes (66, 67), where positioning of the two metal ions can be achieved just as easily by RNA backbone phosphates, proceed through a similar two-metal ion mechanism as that proposed for the 3'  $\rightarrow$  5' exonuclease of pol I (23, 68). In much the same way, perhaps the nucleotidyl transfer mechanism that we present also applies, to some extent, to those ribozymes that are capable of catalyzing the nucleotidyl transfer reaction (69, 70).

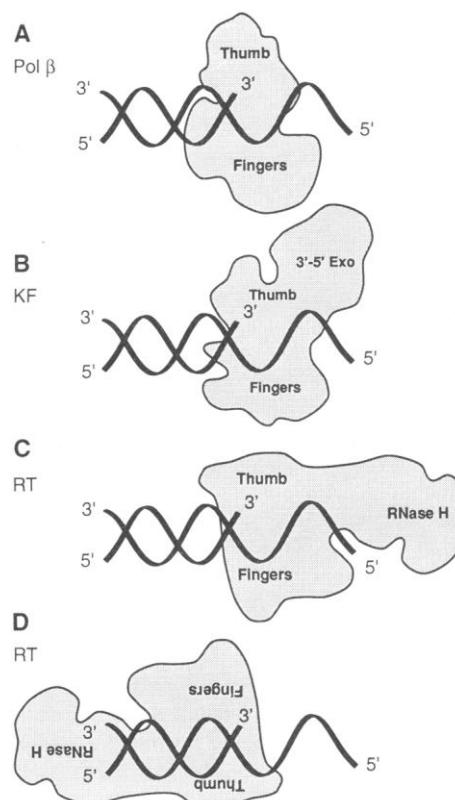
#### Comparisons with other polymerases.

As discussed in the accompanying report (14), the most obvious structural overlap among all four polymerases, KF, RNAP, RT, and pol  $\beta$ , consists of a conserved pair of carboxylic acid side chains located in the palm subdomain. As revealed by the pol  $\beta$  active site, the primary function of these carboxylic acids is to position two catalytically critical divalent metal ions. The ob-

servation that even the most divergent polymerase structures retain this catalytic core is compelling evidence that all polymerases share a common catalytic mechanism (11, 14). We believe that the pol  $\beta$  ternary complex structures presented here are physiologically relevant and that the nucleotidyl transfer mechanism that we propose based on these structures represents a common catalytic mechanism found in all polymerases. In the absence of evidence that suggests otherwise, the next logical argument would be that all polymerases, because they share a common catalytic mechanism, should bind to the DNA (or RNA) template-primer in a fashion very similar—at least with respect to the highly conserved catalytic residues—to that in which pol  $\beta$  has attached itself to the DNA template-primer in our ternary complex (Fig. 7, A to C). All of the polymerase structures discussed have a DNA binding channel that can grasp a stationary rodlike DNA template-primer in one of two general directions, so the only other possibility for binding to the template-primer, provided that there are no gross conformational changes on the part of DNA or protein, is shown for the RT structure (Fig. 7D). In order to generate Fig. 7D, the template-primer in Fig. 7C was held stationary while the RT molecule was rotated by 180° about an axis normal to the plane of the page and passing through the polymerase active site. The mode of DNA binding depicted in Fig. 7D, which is opposite to that found in pol  $\beta$  (anti-pol  $\beta$ ), is the one proposed for KF (41), for RNAP (13), and for RT (11, 12). We are thus faced with a dilemma.

Three main possibilities can be pursued. First, it could be argued that the other three polymerase-DNA models are basically correct, and that the pol  $\beta$ -DNA structures presented here are not physiologically relevant. However, as stated above, there is strong evidence in favor of the physiological relevance of our structures, and we therefore bypass this possibility for the present. A second possibility could be that all four of the polymerase-DNA models are correct, including the pol  $\beta$ -DNA structures presented here, and that pol  $\beta$  uses an entirely different mode of template-primer binding from the other polymerases. However, that pol  $\beta$  has a palm subdomain that is structurally homologous with the other polymerases (14), and that this palm subdomain also contains the highly conserved catalytic residues, suggests otherwise. The two-metal ion mechanism that we propose for the nucleotidyl transfer reaction seems to depend heavily on the positioning of the metal ions by the conserved carboxylic acid residues. Furthermore, the catalytic site appears to be asymmetric in that the two metal sites possess

different geometries and different binding affinities (65). This leads us to the third possibility, which is the one we choose to pursue in the following discussions. The other polymerases must bind to a template-primer in a manner similar to that of pol  $\beta$  so as to conserve the critical, asymmetric geometry of the active site. An inescapable inference, then, is that the other polymerase-DNA models are not correct with respect to the directionality of the template-primer in the binding channel. Because of the complexity of the existing data and interpretations, we address each of the



**Fig. 7.** (A) Schematic drawing showing the position of the template-primer as seen in the pol  $\beta$  ternary complex structure and (B and C) our proposed models for the KF-DNA and RT-DNA complexes, respectively. To generate (B and C) the palms of KF and RT were aligned with the palm subdomain of pol  $\beta$  as viewed in (A). In the text, (A to C) represent the pol  $\beta$ -like mode of binding to the template-primer. To generate (D), which represents the anti-pol  $\beta$  mode of binding to the template-primer, the template-primer in (C) was held stationary, and the RT molecule was rotated by 180° about an axis normal to the plane of the page and running through the RT active site. The template-primer was kept stationary in all four diagrams because that is a reasonable representation of in vivo polymerization, where the DNA molecule is usually very long and immovable in comparison to the polymerase molecule. The active site in all four diagrams is at the 3' terminus of the primer strand, and the view looks down at the palm, which forms the bottom of the template-primer binding channel.

three other polymerases, KF, RNAP, and RT, separately.

DNA pol I of *E. coli* is a 103-kD monomer and can be cleaved by limited proteolysis into a large (68-kD) COOH-terminal fragment [commonly referred to as the Klenow fragment (KF)], and a smaller (35-kD) NH<sub>2</sub>-terminal fragment (1). KF has both polymerase and 3' → 5' exonuclease (editing) activities, whereas the small NH<sub>2</sub>-terminal fragment functions solely as a 5' → 3' exonuclease. The crystal structure of KF alone revealed two distinct domains (10), confirming previous proposals that the polymerase and 3' → 5' exonuclease functions of KF lie on separate, independent folding units. From its position in the pol I amino acid sequence, it was then inferred that the missing NH<sub>2</sub>-terminal 5' → 3' exonuclease domain should be positioned somewhere to the right of the 3' → 5' exonuclease as viewed in Fig. 7B (10). During nick translation, a function intrinsic to pol I, it is proposed that the polymerase, starting at a nick in the DNA and polymerizing in a 5' → 3' direction, works in conjunction with the activities of the 5' → 3' exonuclease so that the net result is simply a translation of the nick along the DNA in a 5' → 3' direction (1). It was this property of pol I that led to proposals that the polymerase domain of pol I bound to the DNA template-primer in a manner not too different from that shown in Fig. 7B, so that the polymerase and the 5' → 3' exonuclease were positioned properly, relative to the DNA nick, to perform their separate activities in conjunction with one another (10). Subsequent KF-DNA models were proposed to show how the polymerase and the 3' → 5' exonuclease of KF could work in conjunction with one another during the DNA editing process (23, 24, 71, 72). Once again, these earlier models require that KF binds to the DNA template-primer in a manner similar to that shown in Fig. 7B. Therefore, all of the earlier pol I-DNA models are in agreement with a pol β-like mode of DNA binding.

The recent crystal structure determination of KF crystallized in the presence of a putative template-primer revealed an unexpected complex in which the KF had bound to the template-primer in neither of the two binding modes discussed above, but instead, in such a way that the DNA lay in a separate, less obvious channel running between the thumb subdomain of the polymerase and the 3' → 5' exonuclease (41). From this structure it was proposed that during polymerization, the KF bent the DNA by 80° so that the DNA entered the larger, more obvious DNA binding channel in a direction opposite to that of all the previously proposed KF-DNA models. We now find ourselves in the awkward position

of presenting arguments in favor of the idea that the initial KF-DNA models were correct, at least with respect to the direction of DNA binding, and that the recent KF-DNA crystal structure (41) might not be physiologically relevant.

Possibly the most compelling reason to believe that the previously reported KF-DNA crystal structure (41) might not be physiologically relevant is that it is not in accord with the nick translation activities of pol I. Other reasons, which concern the nature of the KF-DNA crystals themselves, are (i) the complex crystals were grown under relatively high, nonphysiological salt concentrations; (ii) the template-primer in the crystallization experiments included an unnatural epoxy-ATP; (iii) the template-primer had been unexpectedly modified during crystallization in such a way that the primer strand was longer than the template strand in the crystal structure; and finally, (iv) the complex crystals were isomorphous with the apo KF crystals, which in itself is not a strong argument, but is made stronger by the observation that none of the pol β ternary complex crystals presented here is isomorphous with the apo crystals. Because the 3' → 5' exonuclease domain of KF was shown to have a high affinity for the primer terminus during other attempts at obtaining KF-DNA complex crystals (23), perhaps the recently reported KF-DNA structure (41) is solely an editing complex in which the primer terminus bound to the 3' → 5' exonuclease site, and the leftover "tethered" DNA, guided by crystal packing forces, bound wherever it could under the circumstances. This is similar to the argument presented above that the position of the tethered 8-kD domain in our complex crystals, though ordered in both structures, is probably not physiologically relevant.

Further support for the idea that the KF binds to the template-primer in a manner very similar to that of pol β comes from closer inspection of the DNA binding channels of both enzymes. KF possesses what appears to be the structural equivalent to helices M and N of the pol β thumb subdomain. As described above, helices M and N protrude into the pol β DNA binding channel and interact with the nucleotide substrate (Fig. 3). Though not structurally homologous with helices M and N, helices J and K of KF also protrude into the KF-DNA binding channel in an area near the catalytic carboxylic acids that overlaps quite well with helices M and N of pol β, suggesting a similar role for helices J and K in nucleotide specificity for KF.

RNA polymerase (RNAP) from bacteriophage T7 is a monomer of molecular weight 99-kD, and the crystal structure determination of RNAP revealed that it folds into two distinct domains (13). One domain possesses

a polymerase fold that is highly homologous with the polymerase domain of KF, while a smaller, NH<sub>2</sub>-terminal domain, though not an exonuclease, is located in the same relative position as the 3' → 5' exonuclease domain of KF. That the polymerase domains of RNAP and KF are so structurally homologous was one of the strongest arguments presented in favor of a model (13) wherein RNAP binds to a DNA-RNA duplex in a manner similar to that of the recent KF-DNA model (41). However, as pointed out above, we believe that the recent KF-DNA complex structure might not be physiologically relevant and that the KF, instead, probably utilizes a pol β-like mode of binding to the template-primer. Therefore, we propose that like KF, RNAP also utilizes a pol β-like mode of binding to an RNA-DNA duplex.

Other evidence besides homology with KF was also presented in favor of an anti-pol β mode of template-primer binding for RNAP (13)—in disagreement with our proposals for an RNAP-RNA-DNA complex structure. Results from mutagenesis studies (73, 74) as well as binding studies with proteolytically modified RNAP fragments (75) suggest that the NH<sub>2</sub>-terminal domain of RNAP binds to the nascent RNA of the emerging RNA-DNA duplex (73, 75) and that Gln<sup>748</sup> of RNAP recognizes the -10 and -11 bases of the DNA promoter upon complex formation (74). These observations suggested positioning the RNA-DNA duplex in the template-primer binding channel of RNAP in an anti-pol β fashion (13). Although more work will be required to settle this issue, the structural evidence presented by the pol β ternary complex seems very compelling and interpretations of results from the mutagenesis studies as well as binding studies of proteolytically modified RNAP fragments may require a more careful analysis. As an example of the difficulties involved in interpreting mutagenesis and binding studies in the absence of structural data on a polymerase-DNA complex, mutagenesis studies on pol β implicated Arg<sup>183</sup> as taking part in primer strand recognition upon complex formation (76), and binding studies with pyridoxal 5'-phosphate suggested that the 8-kD domain of pol β formed a part of the nucleotide binding pocket (77). Neither of these conclusions is in agreement with our ternary complex structures, though the results from these studies, when re-evaluated, are not necessarily in disagreement with our structural work [reference 41 in (14)].

Retroviral RTs, responsible for making double-stranded DNA copies of single-stranded RNA viral genomes, had been well studied (78) prior to the discovery that the retrovirus HIV-1 was the cause of AIDS (79-81). This facilitated the study of the

HIV-1 RT because its mechanism of operation is similar to that of many previously studied RTs (82). HIV-1 RT functions as a heterodimer, and the crystal structure determination of RT showed that one monomer of the heterodimer, called p66 because it has a molecular mass of 66 kD, folds into two distinct domains: a typical polymerase domain (palm, fingers, and thumb) and a connected ribonuclease (RNase) H domain (11, 12). The second monomer of the RT heterodimer, p51, is simply a copy of p66 in which the 15-kD RNase H COOH-terminal segment has been proteolytically cleaved. In contrast to the fingers, palm, and thumb of the polymerase domain of p66, the fingers, palm, and thumb of p51 occupy relatively different positions in the crystal structure, resulting in something other than a typical polymerase fold for p51 in the p66-p51 heterodimer (11, 12). Only p66 is shown in Fig. 7C for clarity and to highlight how the RNase H domain lies in approximately the same direction with respect to the polymerase as the 3' → 5' exonuclease domain of KF. The function of the RNase H is to remove the viral RNA template from the RNA-DNA hybrid that results after RNA-directed DNA polymerization (reverse transcription) of the viral genome.

In keeping with our proposals that all polymerases share a common catalytic mechanism, and hence a common template-primer binding mode, we suggest that the polymerase domain of the p66 monomer of RT also binds to a template-primer in a manner similar to that of pol  $\beta$  (Fig. 7C). Some evidence of this exists in that, as with KF and RNAP, RT also seems to possess the structural equivalent to helices M and N of pol  $\beta$ . Though not quite as obvious and pronounced as in the other polymerases, beta strands 12 and 13 of RT do protrude into the template-primer binding channel near the catalytically important carboxylic acid residues in an area that overlaps quite well with helices M and N of pol  $\beta$  and helices J and K of KF, implicating beta strands 12 and 13 of RT as possibly playing a similar role in substrate specificity.

In disagreement with this proposal is the crystal structure of RT complexed with a template-primer, which shows that HIV-1 RT has bound to the template-primer in a manner opposite to that of pol  $\beta$  (12) (Fig. 7D may serve as an approximate representation of that structure). We can present some of the same reservations as to the physiological relevance of the RT-DNA complex as were presented above for the KF-DNA complex: (i) The RT-DNA complex crystals were grown at relatively high, nonphysiological salt concentrations; (ii) the template-primer in the crystallization experiments was odd in that it had only a

single base template overhang; and (iii) the complex crystals were isomorphous with the apo structure crystals. However, unlike KF, there exists strong evidence in favor of the idea that the RT nevertheless binds to a template-primer in the manner depicted in Fig. 7D, regardless of the physiological relevance of the RT-DNA crystal structure. This evidence comes from kinetic studies that show a tight temporal coupling between the polymerase and RNase H activities of RT during the reverse transcription process (83). These results suggest that RT binds to the template-primer in such a way that the RNase domain comes in contact with the RNA template of the emerging RNA-DNA duplex during reverse transcription, which would be possible only if the anti-pol  $\beta$  mode of binding were employed, as shown in Fig. 7D.

Although we have presented strong structural evidence in favor of the idea that RT binds to the template-primer in a manner similar to that of pol  $\beta$ , KF, and RNAP, there seems to be equally strong kinetic evidence (83) in support of an anti-pol  $\beta$  mode of template-primer binding for RT during RNase H coupled polymerization activity. Thus it may be that RT can bind to the template-primer in two different catalytically competent ways; an RNase H independent, pol  $\beta$ -like mode of binding and an RNase H-coupled, anti-pol  $\beta$  mode of binding.

The method by which RT makes double-stranded DNA copies of the single-stranded RNA genome is quite complicated (82), so it is not too difficult to imagine that one enzyme performing so many different functions might employ two modes of binding to a template-primer (Fig. 7, C and D). Although this appears to be in disagreement with our arguments that the asymmetric geometry of the polymerase active site must be conserved, closer inspection of the RT active site reveals that our arguments may nevertheless hold true. As it turns out, the 180° rotation performed on the RT in Fig. 7C to produce Fig. 7D resulted in an active site where Asp<sup>185</sup> and Asp<sup>186</sup> of RT (the equivalent to Asp<sup>190</sup> and Asp<sup>192</sup> in pol  $\beta$ ) have simply switched positions, and Asp<sup>110</sup> (the equivalent to pol  $\beta$ 's Asp<sup>256</sup>) has been replaced by Tyr<sup>183</sup>. Instead of Asp<sup>110</sup>, Tyr<sup>183</sup> is now near the primer 3' terminus for RT in the anti-pol  $\beta$  mode of template-primer binding, suggesting that perhaps Tyr<sup>183</sup> performs some of the same functions that Asp<sup>110</sup> performs when RT is in a pol  $\beta$ -like binding mode. Thus the geometry of the active site seems to be conserved. In support of this idea, the YMDD sequence of RT (Tyr<sup>183</sup>, Met<sup>184</sup>, Asp<sup>185</sup>, Asp<sup>186</sup>) is the most highly conserved amino acid sequence in all known RTs, and mutations to the tyrosine residue

of this segment have shown it to be highly critical for catalysis (84). None of the other polymerases discussed, neither pol  $\beta$ , KF, nor RNAP, has a structural equivalent to Tyr<sup>183</sup> in RT.

Another aspect of the models in Fig. 7, C and D, is that a much smaller part of RT is in contact with duplex DNA in Fig. 7C, possibly suggesting a more distributive mode of synthesis when RT invokes a pol  $\beta$ -like mode of template-primer binding. A more distributive mode of polymerization has been observed for RT when a DNA template is utilized instead of an RNA template (85, 86). All this leads us to propose that RT might use a pol  $\beta$ -like mode of binding for a DNA template, and the reverse, anti-pol  $\beta$  mode of binding for an RNA template, thus making it a reverse transcriptase in another sense of the word. The anti-pol  $\beta$  mode of template-primer binding (Fig. 7D) could be operative during the highly processive, RNase H-coupled, RNA-directed DNA polymerization of the viral minus strand (82), whereas a pol  $\beta$ -like mode of template-primer binding (Fig. 7C) could be employed during the more distributive DNA-directed DNA polymerization of the viral plus strand (82), when RNase H coupled polymerization is no longer required. The most salient feature of this proposal is that RT may distinguish between the two types of DNA substrates involved: the A form of an RNA-DNA hybrid versus the B form of a DNA duplex. For instance, RT might possess the structural equivalents of pol  $\beta$ 's Lys<sup>234</sup>, Tyr<sup>271</sup>, and Arg<sup>283</sup> (which, as discussed above, break up the water structure of the minor groove near the active site) only when RT employs a pol  $\beta$ -like mode of binding. If all DNA-directed DNA polymerization steps require that the water structure in the minor groove of the B-DNA template-primer be disrupted, this would make it unfavorable for RT to bind to a B form substrate in any other way but in a pol  $\beta$ -like fashion.

Although the aforementioned proposals restrict RT from utilizing an anti-pol  $\beta$  mode of binding on B form substrates, they do not necessarily restrict RT from utilizing a pol  $\beta$ -like mode of binding on A-form substrates, which already possess a broad minor groove. It is therefore feasible that other, possibly non-RNase H coupled steps during the replication of the viral genome, such as tRNA primed synthesis of the minus strand strong stop DNA or synthesis of the RNase H insensitive polypurine tract (ppt) (82), might employ a pol  $\beta$ -like mode of template-primer binding for RT as well, despite the fact that an RNA template is utilized during these steps. That it might be possible for RT, under certain circumstances, to employ the same mode of bind-

ing for an RNA template as for a DNA template is supported by kinetic studies showing that pol I, under the appropriate conditions, can utilize an RNA template almost as efficiently as its natural DNA template (87). In the case of the synthesis of minus strand strong stop DNA, the length of the tRNA primer (18 nt) is the same as the distance between the polymerase active site and the RNase H active site (about 18 bp) (83). Accordingly we suggest that perhaps the bulky tRNA molecule attached to the primer strand functions as a steric hindrance to binding at the RNase H active site, forcing RT to use the pol  $\beta$ -like mode of template-primer binding during this step of the cycle. In favor of this argument is the observation that RT polymerization during minus strand strong stop synthesis appears to be more distributive than during reverse transcription of the viral genome (88, 89). Also supporting this idea are results from primer utilization studies showing that tRNA primed synthesis of minus strand strong stop DNA, at least in vitro, does not require a specific tRNA such as human tRNA<sup>3</sup><sub>Lys</sub>, which is utilized by RT in vivo (90).

Further indication of a possible dual mode of template-primer binding by RT comes from many independent studies. Active site titration studies showed that the RT heterodimer possesses a possible second template-primer binding site, but what is most intriguing about these results is that this second binding site only reveals itself after the template strand of the template-primer has been shortened to 16 nt (91). Along the same lines, although a model proposing an RNA-DNA-RNA intermediate for a RT strand transfer mechanism does not suggest a second mode of template-primer binding (92), perhaps some of the kinetic and crosslinking data leading up to that model can also be interpreted as evidence in favor of our arguments. Furthermore, kinetic studies on the individual p66 and p51 monomers of RT showed that both of these monomers are fully capable of catalyzing the nucleotidyl transfer reaction, but only under optimal conditions for each monomer (93). That the optimal template for p66 was RNA and the optimal template for p51 was DNA strengthened the proposal that the RT heterodimer "may be functionally asymmetric with distinct plus and minus strand polymerases" (93). It was further suggested that the p66 monomer of the p66-p51 RT heterodimer was responsible for RNA-directed DNA polymerase functions (reverse transcription), whereas the p51 monomer of the heterodimer was responsible for the DNA-directed DNA polymerization activities of RT (93). However, as discussed above, the crystal structure of the p66-p51 RT heterodimer clearly shows

that the p51 monomer does not possess a polymerase fold, at least while p51 RT is a part of the p66-p51 heterodimer (11, 12). Therefore, our hypothesis that the p66 monomer of the p66-p51 RT heterodimer may act as two different polymerases that happen to share a common active site is not only in agreement with the idea that the RT heterodimer is functionally asymmetric with distinct plus and minus strand polymerases (93), but is in accord with the structural work (11, 12) as well.

Finally, a dual mode of template binding by RT might go a long way toward explaining why drugs such as AZT can specifically target HIV-1 RT in preference to host cell polymerases. Pol  $\beta$ , KF, RNAP, and even RT (when in a pol  $\beta$ -like mode of template-primer binding) all seem to possess structural features near the active site, such as helices M and N of pol  $\beta$ , that may be responsible for nucleotide selectivity. However, there appears to be no structural equivalent to helices M and N of pol  $\beta$  when RT is in an anti-pol  $\beta$  mode of template-primer binding, which might explain why RT is so error prone and more sensitive to AZT-TP inhibition. In support of this idea are kinetic studies that show decreasing  $K_i$  values for AZT-TP in its binding to pol  $\beta$  ( $K_i = 73 \mu\text{M}$ ) and to RT ( $K_i = 0.3 \mu\text{M}$ ) when calf thymus DNA is utilized as a template in both cases, as compared to RT when the native RNA template is utilized ( $K_i = 0.01 \mu\text{M}$ ) (55). The decreasing order of  $K_i$  values (tighter binding) might reflect the decreasing number of structural elements in the active site that can, through steric hindrance, induce nucleotide specificity, the order of structural hindrance being: the pol  $\beta$ 's helices M and N  $\gg$  RT's  $\beta$  strand 12 and 13 (when employing a pol  $\beta$ -like mode of template-primer binding on a DNA template)  $>$  RT when employing an anti-pol  $\beta$  mode of binding on an RNA template. With this in mind, one could imagine how AZT, ddC, and ddI resistant RT mutants might arise in AIDS patients after prolonged treatment with these drugs (94-97). As discussed above, much of the nucleotide binding pocket in pol  $\beta$  is determined by the exact position of the template-primer, so a single mutation that affects how RT binds to the template-primer could in turn affect the nucleotide binding site. Indeed, it has been proposed that many of the drug resistant mutations in RT, because they are so distant from the active site, interact primarily with the template strand (11). Perhaps these RT mutants affect template-primer binding in such a way as to introduce structural elements, similar to helices M and N of pol  $\beta$ , into the nucleotide binding pocket of RT, possibly causing greater selectivity and more resistance to these anti-HIV drugs.

## REFERENCES AND NOTES

1. A. Kornberg and T. A. Baker, *DNA Replication* (Freeman, New York, ed. 3, 1991).
2. R. Yarchoan, H. Mitsuya, C. E. Myers, S. Broder, *N. Engl. J. Med.* **321**, 726 (1989).
3. R. F. Schinazi, *Perspect. Drug Discovery Design* **1**, 151 (1993).
4. H. Mitsuya *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2033 (1987).
5. P. A. Furman *et al.*, *ibid.* **83**, 8333 (1986).
6. ddI is the deamination product of ddA; it is thought that ddIMP, in vivo, is converted to ddAMP by adenylosuccinate synthetase-lyase, which is then converted to ddATP by further phosphorylation [M. A. Johnson *et al.*, *J. Biol. Chem.* **263**, 15354 (1988)].
7. M. A. Waqer, M. J. Evans, K. F. Manly, R. G. Hughes, J. A. Huberman, *J. Cell. Physiol.* **121**, 402 (1984).
8. J.-P. Sommadossi, R. Carlisle, Z. Zhou, *Mol. Pharmacol.* **36**, 9 (1989).
9. Y. C. Cheng, W. Y. Gao, C. H. Chen, M. Vazquez-Padua, M. C. Starnes, *Annals N. Y. Acad. Sci.* **616**, 217 (1990).
10. D. L. Ollis, P. Brick, R. Hamlin, N. G. Xoung, T. A. Steitz, *Nature* **313**, 762 (1985).
11. L. A. Kohlstaedt, J. Wang, J. M. Friedman, P. A. Rice, T. A. Steitz, *Science* **256**, 1783 (1992).
12. A. Jacobo-Molina *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6320 (1993).
13. R. Sousa, Y. J. Chung, J. P. Rose, B. C. Wang, *Nature* **364**, 593 (1993).
14. M. R. Sawaya, H. Pelletier, A. Kumar, S. H. Wilson, J. Kraut, *Science* **260**, 1930 (1994).
15. M. Delarue, O. Pooh, N. Tordo, D. Moras, P. Argos, *Protein Eng.* **3**, 461 (1990).
16. S. Linn, *Cell* **66**, 185 (1991).
17. T. M. Jenkins, J. K. Saxena, A. Kumar, S. H. Wilson, E. J. Ackerman, *Science* **258**, 475 (1992).
18. D. C. Rein, A. J. Recupero, M. P. Reed, R. R. Meyer, *The Eukaryotic Nucleus: Molecular Biochemistry and Macromolecular Assembly*, P. R. Strauss and S. H. Wilson, Eds. (Telford, Caldwell, NJ, 1990), vol. 1, chap. 5.
19. S. H. Wilson, *ibid.*, chap. 8.
20. Certain aspects of the aging process, as well as some types of cancer, may be linked to a breakdown in the DNA repair mechanisms of the cell and hence to pol  $\beta$  in particular. Supporting this view, deleterious mutations in the gene encoding for pol  $\beta$  have been observed in blood samples from patients with Werner syndrome (WS), a rare autosomal recessive disorder characterized by a high DNA mutation rate and the appearance of premature aging [Y. Sadakane, K. Maeda, Y. Kuroda, K. Hori, *Biochem. Biophys. Res. Commun.* **200**, 219 (1994)] and also in tumor samples from patients with colorectal cancer [L. Wang, U. Patel, L. Ghosh, S. Banerjee, *Cancer Res.* **52**, 4824 (1992)].
21. S. S. Carroll and S. J. Benkovic, *Chem. Rev.* **90**, 1291 (1990).
22. R. Green and D. Korn, *J. Biol. Chem.* **245**, 254 (1970). Pol  $\beta$  is so efficient at incorporating mismatched base pairs that it was originally questioned whether pol  $\beta$  utilizes a template at all during DNA polymerization. Fueling the debate was the observation that the only other polymerase sharing sequence similarity with pol  $\beta$ —terminal deoxynucleotidyl-transferase (TdT)—is not a template-directed polymerase.
23. P. S. Freemont, J. M. Friedman, L. S. Beese, M. R. Sanderson, T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8924 (1988).
24. L. S. Beese and T. A. Steitz, *EMBO J.* **10**, 25 (1991).
25. B. Z. Zmudzka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5106 (1986).
26. A. Kumar *et al.*, *J. Biol. Chem.* **265**, 2124 (1990).
27. The oligonucleotides that were to serve as template and primer, 5'-AAAGGGCGCCG-3' and 5'-CGGCGC-3', respectively, were purchased (Integrated DNA Technologies, Inc., Coralville, IA) and stored at  $-20^{\circ}\text{C}$ . Occasional DNA samples

- that had a slight yellow hue and were difficult to dissolve in the buffer solution inevitably failed to produce complex crystals. We therefore recommend that a reversed-phase cartridge, or other DNA purification techniques, be employed before the crystallization experiments.
28. A more common method for annealing the DNA template-primer, in which samples are heated to 90° for 3 minutes then allowed to cool slowly to room temperature, was attempted in order to obtain better crystals, but this method resulted in crystals with no noticeable improvements in diffraction power.
  29. The ddCTP was purchased (Sigma) as 4 μmol samples and stored at -20°C. The H<sub>2</sub>O used throughout all crystallization procedures had been deionized, then distilled, prior to use.
  30. Crystals were grown at room temperature in MVD-24 sitting drop trays (Charles Supper Co.), which were subsequently sealed with clear packaging tape (Manco, purchased from Sears) to allow vapor diffusion between the reservoir solution [7 to 9 percent (w/v) PEG 3350, 0.1 M MES, 75 mM lithium sulfate] and the crystallization medium, which was made by mixing 20 μl of the reservoir solution with 20 μl of the protein-DNA-ddCTP sample.
  31. AZT-TP was provided by R. F. Schinazi, Emory University, Georgia.
  32. A partial (31-kD) structure was determined by molecular replacement techniques, then refined, in a manner very similar to that described in the text for the ternary complex structures. It was difficult to interpret  $F_o - F_c$ ,  $\alpha_c$ , difference maps at this resolution (4 Å), but what little positive electron density that was observed in the maps was located adjacent to the fingers subdomain of pol β and was attributed to the missing 8-kD domain. No further work has been done on this structure.
  33. L. S. Beese, J. M. Friedman, T. A. Steitz, *Biochemistry* **32**, 14095 (1993).
  34. D. N. SenGupta *et al.*, *Biochem. Biophys. Res. Commun.* **136**, 341 (1986).
  35. J. Abbotts *et al.*, *Biochemistry* **27**, 901 (1988).
  36. A. T. Brunger, *XPLOR Manual* (Yale Univ. Press, New Haven, CT, version 3.1, 1992).
  37. A. T. Brunger, *Acta Crystallogr.* **A46**, 46 (1990).
  38. The greater difficulty in obtaining clear translation solutions for the  $P2_1$  structure could be attributed to the fact that the  $P2_1$  structure contains two molecules per asymmetric unit instead of one, as is the case for the  $P6_1$  structure. Also, diffraction data on the  $P2_1$  crystals were limited to a lower resolution than diffraction data on the  $P6_1$  crystals. We were originally unable to obtain clear translation solutions for the  $P2_1$  crystals from a low resolution (4.0 Å) data set, and the molecular replacement structure determination of the  $P2_1$  crystals had to await the growth of slightly better crystals and the somewhat higher resolution data (3.6 Å) that we present (Table 1).
  39. D. E. Tronrud, L. F. TenEyck, B. W. Matthews, *Acta Crystallogr.* **A43**, 489 (1987).
  40. R. Lavery and H. Sklenar, *J. Biomol. Struct. Dynam.* **6**, 655 (1989).
  41. L. S. Beese, V. Derbyshire, T. A. Steitz, *Science* **260**, 352 (1993).
  42. R. Prasad, W. A. Beard, S. H. Wilson, *J. Biol. Chem.*, in press.
  43. R. K. Singhal and S. H. Wilson, *ibid.* **268**, 15906 (1993).
  44. A. Kumar, J. Abbotts, E. M. Karawya, S. H. Wilson, *Biochemistry* **29**, 7156 (1990).
  45. J. Warwicker, D. Ollis, F. M. Richards, T. A. Steitz, *J. Mol. Biol.* **186**, 645 (1985).
  46. B. N. Conner, T. Takano, S. Tanaka, K. Itakura, R. E. Dickerson, *Nature* **295**, 294 (1982).
  47. H. R. Drew and R. E. Dickerson, *J. Mol. Biol.* **151**, 535 (1981).
  48. K. Tanabe, E. W. Bohn, S. H. Wilson, *Biochemistry* **18**, 3401 (1979).
  49. W. R. McClure, T. M. Jovin, *J. Biol. Chem.* **250**, 4073 (1975).
  50. C. Majumdar, J. Abbotts, S. Broder, S. H. Wilson, *ibid.* **263**, 15657 (1988).
  51. R. D. Kuchta, V. Mizrahi, P. A. Benkovic, K. A. Johnson, S. J. Benkovic, *Biochemistry* **26**, 8410 (1987).
  52. J. E. Reardon and W. H. Miller, *J. Biol. Chem.* **265**, 20302 (1990).
  53. Interaction of Gly<sup>274</sup> with ddCTP is noteworthy because Gly<sup>274</sup> is at the COOH-terminus of helix M and, as is mentioned earlier, forms a cis-peptide with Ser<sup>275</sup> of helix N.
  54. J. H. Van de Sande, P. C. Loewen, H. G. Khorrana, *J. Biol. Chem.* **247**, 6140 (1972).
  55. M. H. St. Clair *et al.*, *Antimicrob. Agents Chemother.* **31**, 1972 (1987).
  56. J. E. Reardon, *Biochemistry* **31**, 4473 (1992).
  57. T. Date, S. Yamamoto, K. Tanihara, Y. Nishimoto, A. Matsukage, *ibid.* **30**, 5286 (1991).
  58. A. H. Polesky, T. A. Steitz, N. D. F. Grindley, C. M. Joyce, *J. Biol. Chem.* **265**, 14579 (1990).
  59. A. H. Polesky, M. E. Dahlberg, S. J. Benkovic, N. D. F. Grindley, C. M. Joyce, *ibid.* **267**, 8417 (1992).
  60. B. A. Larder, D. J. M. Purifoy, K. L. Powell, G. Darby, *Nature* **327**, 716 (1987).
  61. P. M. J. Burgers and F. Eckstein, *J. Biol. Chem.* **254**, 6889 (1979).
  62. V. Derbyshire *et al.*, *Science* **240**, 199 (1988).
  63. V. Derbyshire, N. D. F. Grindley, C. M. Joyce, *EMBO J.* **10**, 17 (1991).
  64. A. P. Gupta and S. J. Benkovic, *Biochemistry* **23**, 5874 (1984).
  65. In the refined  $P6_1$  ternary complex structure, the temperature factor for the Mg<sup>2+</sup> ion in site B (79 Å<sup>2</sup>) was much greater than that of the Mg<sup>2+</sup> ion in site A (6 Å<sup>2</sup>). Because the electron density in site B could also have been interpreted as a water molecule, we had to rely on local geometry to distinguish the Mg<sup>2+</sup> ion in site B from a water; in particular, a short coordinate bond of 2.0 Å to the oxygen of Asp<sup>190</sup> (Table 3) is characteristic of a metal ion. The low occupancy of the metal ion in site B may be attributed to the fact that one of its ligands, the 3' oxygen of the primer terminus, is missing in the crystal structure. Because we have consistently observed [(14), H. Pelletier and M. R. Sawaya, unpublished results] that only the metal ion in site B is present in 31-kD pol β crystals that have been soaked in the presence of MnCl<sub>2</sub>, we believe that the second metal ion (site A) most likely enters the active site with the nucleotide substrate as a β,γ-bidentate and exists as M<sup>2+</sup>-PP<sub>i</sub>.
  66. T. R. Cech, *Science* **236**, 1532 (1987).
  67. A. M. Pyle, *ibid.* **261**, 709 (1993).
  68. T. A. Steitz and J. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498 (1993).
  69. M. D. Been and T. R. Cech, *Science* **239**, 1412 (1988).
  70. D. P. Bartel and J. W. Szostak, *ibid.* **261**, 1411 (1993).
  71. C. M. Joyce and T. A. Steitz, *Trends Biochem. Sci.* **12**, 288 (1987).
  72. T. A. Steitz, L. Beese, P. S. Freemont, J. M. Friedman, M. R. Sanderson, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 465 (1987).
  73. D. Patra, E. M. Lafer, R. Sousa, *J. Mol. Biol.* **224**, 307 (1992).
  74. C. A. Raskin, G. Diaz, K. Joho, W. T. McAllister, *ibid.* **228**, 506 (1992).
  75. D. K. Muller, C. T. Martin, J. E. Coleman, *Biochemistry* **27**, 5763 (1988).
  76. T. Date *et al.*, *ibid.* **29**, 5027 (1990).
  77. A. Basu, P. Kedar, S. H. Wilson, M. J. Modak, *ibid.* **28**, 6305 (1989).
  78. E. Gilboa, S. W. Mitra, S. Goff, D. Baltimore, *Cell* **18**, 93 (1979).
  79. F. Barré-Sinoussi *et al.*, *Science* **220**, 868 (1983).
  80. M. Popovic, M. G. Sarngadharan, E. Read, R. C. Gallo, *ibid.* **224**, 497 (1984).
  81. R. C. Gallo *et al.*, *ibid.*, p. 500.
  82. S. P. Goff, *J. Acquired Immune Deficiency Syndrome* **3**, 817 (1990).
  83. V. Gopalakrishnan, J. A. Peliska, S. J. Benkovic, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10763 (1992).
  84. S. A. Jablonski and C. D. Morrow, *J. Virol.* **67**, 373 (1993).
  85. H. E. Huber, J. M. McCoy, J. S. Seehra, C. C. Richardson, *J. Biol. Chem.* **264**, 4669 (1989).
  86. J. E. Reardon, *ibid.* **268**, 8743 (1993).
  87. M. Ricchetti and H. Buc, *EMBO J.* **12**, 387 (1993).
  88. W. A. Haseltine, D. G. Kleid, A. Panet, E. Rothenberg, D. Baltimore, *J. Mol. Biol.* **106**, 109 (1976).
  89. L. I. Lobel and S. P. Goff, *J. Virol.* **53**, 447 (1985).
  90. L. A. Kohlstaedt and T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9652 (1992).
  91. W. A. Beard and S. H. Wilson, *Biochemistry* **32**, 9745 (1993).
  92. J. A. Peliska and S. J. Benkovic, *Science* **258**, 1112 (1992).
  93. R. L. Thimmig and C. S. McHenry, *J. Biol. Chem.* **268**, 16528 (1993).
  94. B. A. Larder, G. Darby, D. D. Richman, *Science* **243**, 1731 (1989).
  95. B. A. Larder and S. D. Kemp, *ibid.* **246**, 1155 (1989).
  96. M. H. St. Clair *et al.*, *ibid.* **253**, 1557 (1991).
  97. J. E. Fitzgibbon *et al.*, *Antimicrob. Agents Chemother.* **36**, 153 (1992).
  98. R. Hamlin, *Methods Enzymol.* **114**, 416 (1985).
  99. A. J. Howard, C. Nielsen, N. H. Xuong, *ibid.*, p. 452 (1985).
  100. P. J. Kraulis, *J. Appl. Cryst.* **24**, 946 (1991).
  101. Supported in part by NIH grants GM10928 and CA17374 (J.K.), NIH grant ES06839, R. A. Welch Foundation grant H-1265 (S.H.W.), and a grant of computing time from the San Diego Supercomputer Center. We thank the personnel of the N. H. Xuong Laboratory for aid in data collection; C. Nielsen, N. Nguyen, D. Sullivan, V. Ashford, and W. Wolfe for technical assistance in protein preparation. Full coordinates for both ternary complex structures are available from the Brookhaven Protein Data Bank and are designated 1bpf and 1bpg for the  $P6_1$  and  $P2_1$  structures, respectively.

28 February 1994; accepted 10 May 1994