JOURNAL OF BACTERIOLOGY, Nov. 1995, p. 6321–6329 Vol. 177, No. 22
0021-9193/95/$04.00+0

MINIREVIEW

Polymerase Structures and Function: Variations on a Theme?

CATHERINE M. JOYCE[1]* AND THOMAS A. STEITZ[1,2]

Department of Molecular Biophysics and Biochemistry[1] and Howard Hughes Medical Institute,[2] Yale University, New Haven, Connecticut 06520

In recent years several lines of evidence have led to the proposal that all nucleic acid polymerases show fundamental similarities in structure and the mechanism of catalysis. First, protein sequence alignments suggested that all polymerases contain the same group of important side chains, which include two or, more commonly, three carboxylates. Second, crystallographic studies of four polymerases showed the polymerase domain of each enzyme folded to form a U-shaped cleft with the conserved carboxylates located in similar positions in the four structures. Finally, site-directed mutagenesis experiments have demonstrated that the conserved residues, particularly the carboxylates, play an important role in the polymerase reaction.

The apparent similarity in the core polymerase structures cuts across the division of polymerases into classes based on whether the template and the synthesized strand are DNA or RNA. The implication is therefore that the basic mechanism of phosphoryl transfer is the same throughout the polymerase superfamily, although additional features must be present to account for the use of deoxyribo- or ribo-substrates in the four classes of polymerase. (This idea is consistent with the observation that polymerases can often be induced, by changing the reaction conditions, to use the "wrong" substrate.) A second obvious distinction between polymerase classes is the presence of additional functions: examples are the proofreading 3′-5′ exonuclease of many DNA-dependent DNA polymerases, the 5′-3′ exonuclease of some bacterial DNA polymerases, the RNase H activity of reverse transcriptases, and the binding of replicative polymerases to accessory proteins. However, there is reason to believe that these additional functions do not affect the basic properties of the polymerase core, since the known polymerase structures indicate a modular arrangement, with additional enzymatic activities present on independent structural domains.

The four polymerase structures currently known represent four distinct polymerase families and three polymerase classes. The Klenow fragment of *Escherichia coli* DNA polymerase I (Pol I), a DNA-dependent DNA polymerase, is a monomer of 68 kDa which has, in addition to the polymerase domain, a separate structural domain containing the 3′-5′ proofreading exonuclease (32). A second DNA-dependent DNA polymerase structure, that of rat DNA polymerase β (Pol β), has recently been reported (14, 40). This 39-kDa protein contains a separate 8-kDa N-terminal domain with single-stranded DNA binding and deoxyribose phosphate excision activities (27, 28, 30a) which has been proteolytically removed in some of the crystal forms. The reverse transcriptase from human immunodeficiency virus type 1 (HIV-1) is an example of an RNA-dependent DNA polymerase (21, 26). This enzyme is a heterodimer having one enzymatically active 66-kDa subunit containing polymerase and RNase H activities on separate domains. The second subunit (51 kDa), though derived from the same amino acid sequence as the polymerase domain, is folded so as to preclude enzymatic activity (26); instead, this subunit may interact with the tRNA molecule that is used as a primer for reverse transcription. The fourth structure, that of the 99-kDa monomeric RNA polymerase from bacteriophage T7 (41), represents the class of DNA-dependent RNA polymerases. In addition to the polymerase domain, this molecule contains a separate N-terminal domain of about 300 amino acids that is probably involved in functions specific to RNA synthesis.

For three of these polymerases (Klenow fragment, Pol β, and HIV-1 reverse transcriptase), cocrystal data provide information on the binding of DNA to the polymerase domain (4, 21, 34). In each case, the crystallographic data lead to a model entirely consistent with existing biochemical and molecular genetic data for that enzyme system. Nevertheless, others have suggested (1, 34) that some of the cocrystal structures are not relevant because of an apparent disagreement between the conclusions reached from these structures. We present here our contrasting view that this disagreement is the result of an inappropriate alignment of nonhomologous structures. When the structures are appropriately aligned, the expected similarities in substrate binding and catalysis at the polymerase site are preserved.

In this article we shall ignore the other activities present in some polymerases in order to focus on the current understanding of how the polymerase domain itself is constructed and how catalysis of the polymerase reaction is achieved.

PROTEIN SEQUENCE ALIGNMENTS

Figure 1 shows an alignment of polymerase sequences based on that originally proposed by Delarue et al. (16). Although there is very little overall resemblance among sequences from different classes of polymerases or from the different families (Pol I, Pol α, and Pol β) within the class of DNA-dependent DNA polymerases, Delarue et al. were able to align all polymerase sequences so as to obtain two conserved sequence motifs (A and C) containing carboxylate residues. Another motif (B), containing an invariant lysine, was present in all DNA-dependent polymerases, and there are, of course, several additional motifs (not shown) which are confined to particular families or classes of polymerases. The proposed alignment is supported by mutational data showing the importance of many of the conserved residues (for a recent review, see reference

* Corresponding author. Mailing address: Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208114, 266 Whitney Ave., New Haven, CT 06520-8114. Phone: (203) 432-8992. Fax: (203) 432-9782. Electronic mail address: cjoyce@minerva.cis.yale.edu.

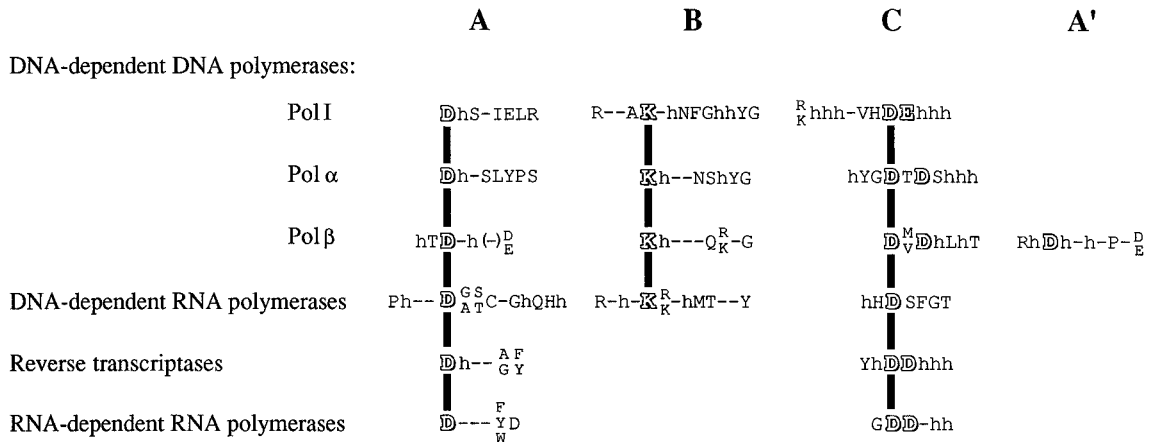


FIG. 1. Alignment of the major conserved sequence regions of the polymerase families. The listed motifs are based on published compilations for the Pol I (11, 16), Pol α (11, 46), and Pol β (19, 20) families, for the single-subunit DNA-dependent RNA polymerases (31), and for RNA-dependent polymerases (36). Invariant and highly conserved amino acids are indicated by white outlined letters. Positions that are almost invariably occupied by a hydrophobic amino acid are indicated (h). Hyphens denote nonconserved positions; parentheses are used to indicate length variations within a motif. For clarity, many of the motifs that are conserved within individual families are not shown. The black lines indicate the alignment proposed by Delarue et al. (16) which gives two motifs, A and C (labeled at the top of the figure), common to the entire polymerase family, containing two invariant aspartates and another highly conserved acidic residue (in motif C). Motif B, containing an invariant lysine, is common to DNA-dependent polymerases (16, 19). For the Pol β family, A′ indicates the sequence motif (taken from sequences compiled in reference 20) containing Asp-256, which was identified from the Pol β protein structure as the third active-site carboxylate (14). Note that the small number of sequences in the Pol β family makes it particularly hard to identify with confidence the important conserved amino acids.

23). Moreover, in three of the four polymerase structures, the motif A and C carboxylates form a cluster at the polymerase active site. The exception is the Pol β structure; of the three carboxylates at the active site, two (Asp-190 and Asp-192) are derived from motif C, but the third (Asp-256) is located C terminal to motif C (14), not N terminal (Asp-17) as suggested by the alignment of Delarue et al. (16). The identity of the third active-site carboxylate of Pol β has been confirmed by mutagenesis (14). Thus, the connectivity of the Pol β structure is sufficiently different from that of the other three polymerases that the active-site residues appear in a different order on the protein sequence, arguing against a common evolutionary origin for Pol β and the other polymerases whose structures are known. Moreover, the Pol β precedent indicates that one should be cautious in making inferences from the sequence alignments shown in Fig. 1 in the absence of structural information for the particular polymerase family under consideration, since in most cases the degree of sequence identity, when polymerases of different families are compared, is insufficient to allow an unambiguous alignment in the absence of other data.

POLYMERASE DOMAIN STRUCTURES

In each of the four published polymerase structures, the polymerase domain folds so as to form a large cleft. The shape of the domain has been compared with that of a half-open right hand (32), and the subdomains that define the base and walls of the cleft have been described as "palm," "fingers," and "thumb," respectively (26) (Fig. 2). Detailed comparison shows that the four structures differ in their relatedness to one another. Klenow fragment and T7 RNA polymerase presumably must share a common ancestor, since their entire polymerase domains are extremely similar, with identical connectivities and arrangements of secondary structure elements (41). In reverse transcriptase, the palm subdomain appears to be homologous to that of Klenow fragment or T7 RNA polymerase, though the fingers and thumb are quite different and the connectivity of the three subdomains differs from that of Klenow fragment (26). An alternative view is that there are a limited number of ways of constructing a palm subdomain so that the resemblance between the reverse transcriptase palm and that of Klenow fragment is purely coincidental and does not reflect a homologous relationship. This interpretation cannot be ruled out at present, though it should be possible to assess its validity as more polymerase structures are determined.

By contrast, the structural data indicate that Pol β is not homologous to Klenow fragment, reverse transcriptase, or T7 RNA polymerase, since even the palm subdomain of Pol β differs in some important respects from its counterparts in the other three structures (14, 40). Despite the lack of a common evolutionary origin for all four polymerases, the palm, fingers, and thumb subdomains appear to serve analogous functions in the four structures. The palm subdomain contains the catalytic site and must therefore be responsible for binding the primer terminus and the α-phosphate of the incoming deoxynucleoside triphosphate (dNTP). The thumb subdomain interacts with the template-primer stem upstream of the site of synthesis, and the fingers subdomain binds the single-stranded template across from and beyond the site of synthesis.

In this section we shall first describe the conclusions that have been drawn from the three related structures (Klenow fragment, reverse transcriptase, and T7 RNA polymerase). Then we shall describe the ways in which the nonhomologous Pol β structure has added to our insights into polymerase structure and mechanism. By showing that the polymerase active-site structure can accommodate greater variations than were previously suspected, the Pol β structure allows us to draw inferences about what are the "bare essentials" of a polymerase active site.

The palm subdomain–location of the catalytic site. As described above, the palm subdomain, which consists largely of the β-sheet that forms the base of the polymerase cleft and the two long helices that pack against this sheet, is the best-conserved portion of the Klenow fragment, reverse transcriptase, and T7 RNA polymerase structures. Four β-strands and two α-helices are essentially superimposable (26, 41, 44), as illustrated in Fig. 3 for Klenow fragment and reverse transcriptase.

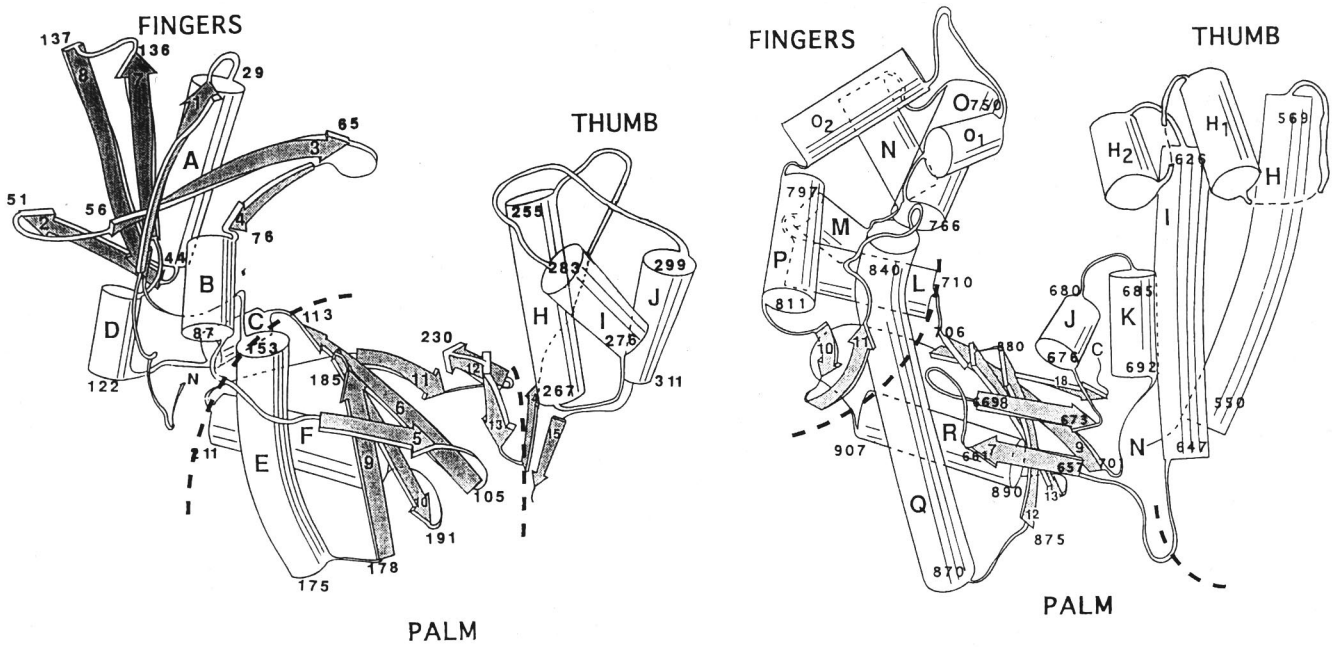


FIG. 2. Schematic representation of the polymerase domain of HIV-1 reverse transcriptase (left) and Klenow fragment (right) with α-helices shown as cylinders and β-strands as arrows (figure reprinted from *Cold Spring Harbor Symposia on Quantitative Biology* [44] with permission of the publisher). The molecules are positioned with their palm subdomains oriented identically. The amino acid sequence numbers give approximate indications of the boundaries of secondary structure features. Note that further refinement of the Klenow fragment structure (5a) has allowed identification of four additional helices (H_1, H_2, O_1, and O_2) that were not apparent on the original 3.3 Å (ca. 0.3 nm) structure (32).

This superposition results in virtually identical positioning of three carboxylate side chains (Asp-705, Asp-882, and Glu-883 of Klenow fragment and Asp-110, Asp-185, and Asp-186 of reverse transcriptase), which correspond to the single carboxylate of motif A and the pair of adjacent carboxylates of motif C (Fig. 1) (16). The corresponding carboxylates of T7 RNA polymerase (Asp-537 and Asp-812) are also reported to be in similar locations (41).

The presence of highly conserved carboxylates in the protein sequence alignments and the essentially identical positioning of the carboxylates in the Klenow fragment, reverse transcriptase, and T7 RNA polymerase structures implies that they may be important active-site side chains. This expectation has been borne out by site-directed mutagenesis experiments on these three polymerases, showing that substitution of the carboxylates caused a dramatic decrease in polymerase activity in every case (8, 9, 29, 30, 33, 37, 38). Mutagenesis experiments have also implicated other polar residues from the same general area of the polymerase cleft in various aspects of polymerase function (reference 23 and references cited therein). In Klenow fragment and T7 RNA polymerase, amino acids that appear to be important in the polymerase reaction include the conserved residues of sequence motif B, particularly the invariant lysine (3, 8, 33). These residues are located on a long helix in the fingers subdomain (helix O in Klenow fragment [Fig. 2]), with the conserved side chains exposed on the surface of the polymerase cleft. Thus, the polymerase catalytic site may extend from the carboxylate region of the palm into the neighboring portion of the fingers subdomain, consistent with evidence (see below) that the dNTP substrate interacts with residues from both the palm and fingers subdomains.

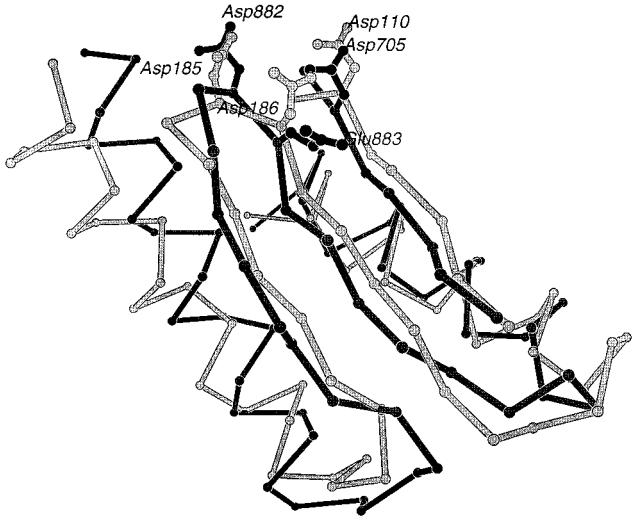


FIG. 3. Superposition of a portion of the palm subdomains of Klenow fragment (in darker shading) and HIV-1 reverse transcriptase, showing the α-carbon backbone and the three conserved carboxylate residues (figure reprinted from *Cold Spring Harbor Symposia on Quantitative Biology* [44] with permission of the publisher).

Fingers and thumbs–polynucleotide binding. In contrast with the palm subdomains, very little structural similarity is seen when the fingers and thumb subdomains of reverse transcriptase are compared with those of Klenow fragment or T7 RNA polymerase, although the available evidence suggests that these subdomains may serve the same functions in the three enzymes. The thumb subdomains show a superficial resemblance in that they all have largely helical structures. By contrast, these three enzymes exemplify two distinct families of structures for the fingers subdomain (Fig. 2). Klenow fragment and T7 RNA polymerase both have predominantly α-helical fingers subdomains with closely related structures (32, 41). In HIV-1 reverse transcriptase, the fingers subdomain has a completely different structure containing both α-helix and β-sheet

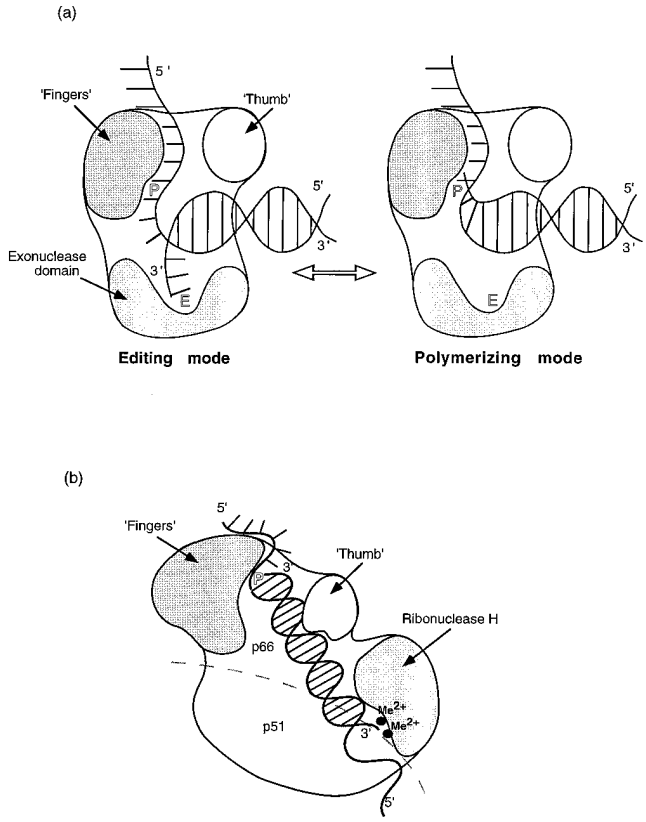


FIG. 4. Schematic representation of the location of DNA bound to Klenow fragment (a) and HIV-1 reverse transcriptase (b) (figure reprinted from *Current Opinion in Structural Biology* [42] with permission of the publisher). In both enzymes, "P" indicates the location of the polymerase active site. For Klenow fragment, "E" indicates the 3′-5′ exonuclease active site; in reverse transcriptase, the two metal ions (Me^{2+}) mark the location of the RNase H active site. Note that the direction of DNA synthesis relative to the conserved polymerase catalytic subdomain appears likely to be the same for both polymerases, although the way in which the DNA duplex approaches the active site is different.

(26). Protein sequence comparisons reinforce the structural conclusions and suggest that the two types of fingers structures may reflect the template preference of a polymerase, in that the B sequence motif of Delarue et al. (16), located within the fingers subdomain, is found only in DNA-dependent polymerases.

The roles of the fingers and thumb subdomains in binding the polynucleotide substrate have been inferred from DNA cocrystal structures and from model building. In the complex of HIV-1 reverse transcriptase with a duplex DNA oligonucleotide (21), the duplex DNA enters the cleft from the end closest to the RNase H domain, following a path very similar to that previously derived by model building a DNA-RNA hybrid onto the native structure (26), as shown schematically in Fig. 4b. A cocrystal structure of Klenow fragment with duplex DNA has also been determined (4). Although the Klenow fragment complex has the DNA primer terminus at the 3′-5′ exonuclease (editing) site, the interactions seen between the protein and the duplex portion of the DNA may well be relevant to the corresponding complex in which the primer terminus is at the polymerase site (Fig. 4a). This structure shows duplex DNA bound in a cleft formed between the polymerase and exonuclease domains of Klenow fragment. In order to position the primer terminus in the polymerase site in a manner similar to that seen in the reverse transcriptase complex, the duplex DNA must make a sharp turn as it enters the polymerase cleft. In both cocrystals the thumb subdomain is seen to contact the minor groove of the primer-template duplex upstream of the site of synthesis (4, 21). Moreover, the binding of DNA in the cocrystal results in movement of the thumb from its position in the native structure (2, 4, 21), implying that there is some flexibility in the attachment of this subdomain to the rest of the polymerase molecule. This apparent flexibility could be important in allowing access of the template-primer to the binding site and in translocation of the elongated DNA to allow another cycle of dNTP addition.

Although neither cocrystal provides information on the location of the uncopied template strand beyond the site of nucleotide addition, model building by extending the template strand from the observed duplex DNA positions necessarily places the uncopied template in contact with the fingers subdomain (4, 21, 26). Thus, an important function of the fingers subdomain could be to bind and orient the template strand. As mentioned above, the sequence and structural data together suggest that this subdomain may provide template specificity, with the helical Klenow fragment-like structure being optimal for a DNA template, the mixed α-β reverse transcriptase structure being appropriate for binding either RNA or DNA (as required during retroviral replication), and, perhaps, yet another fingers structure in those polymerases that use an RNA template exclusively. However, more polymerase structures will need to be determined before the validity of this generalization can be assessed.

POLYMERASE REACTION MECHANISM

The studies of Klenow fragment, reverse transcriptase, and T7 RNA polymerase suggested a mechanism for the polymerase reaction (5, 42); further evidence in support of this mechanism has recently been provided by structural studies on a ternary complex of Pol β with both its substrates (DNA and dNTP) (34). As described above, the polymerase active site is defined by a cluster of conserved carboxylates and other polar residues at the base of the polymerase cleft. The carboxylates, in particular, are crucially important for catalysis of the polymerase reaction, a phosphoryl transfer reaction involving nucleophilic attack by the 3′ hydroxyl of the primer terminus on the dNTP α-phosphate, with release of PP_i. In the proposed mechanism (Fig. 5), these carboxylates serve to anchor a pair of divalent metal ions which then play the major role in catalysis; one divalent metal ion (shown as Mg^{2+} ion 1) promotes the deprotonation of the 3′ hydroxyl of the primer strand, while the other (Mg^{2+} ion 2) facilitates the formation of the pentacovalent transition state at the α-phosphate of the dNTP and the departure of the PP_i leaving group. Two-metal-ion catalysis of this type may, in fact, be a recurrent theme in many phosphoryl transfer reactions (45); there is evidence for this type of mechanism at the 3′-5′ exonuclease active site of Klenow fragment (5, 18), and in alkaline phosphatase (25), and the same mechanism has been proposed for ribonuclease H (15, 47) and for ribozymes (18, 35, 45). Support for this two-metal-ion mechanism for polymerases originally came from crystallographic experiments with Klenow fragment that showed binding of divalent metal ions in the region of Asp-705 and Asp-882 (5a). More recently, the Pol β ternary complex structure (34) showed clearly the coordination of two metal ions, separated by about 4 Å (0.4 nm). This is the same spacing observed for the pairs of metal ions in the 3′-5′ exonuclease, alkaline phosphatase, and ribonuclease H structures (5, 15, 25). The binding of the dNTP α-phosphate bridging the two metal ion positions, and the positioning of the primer terminus

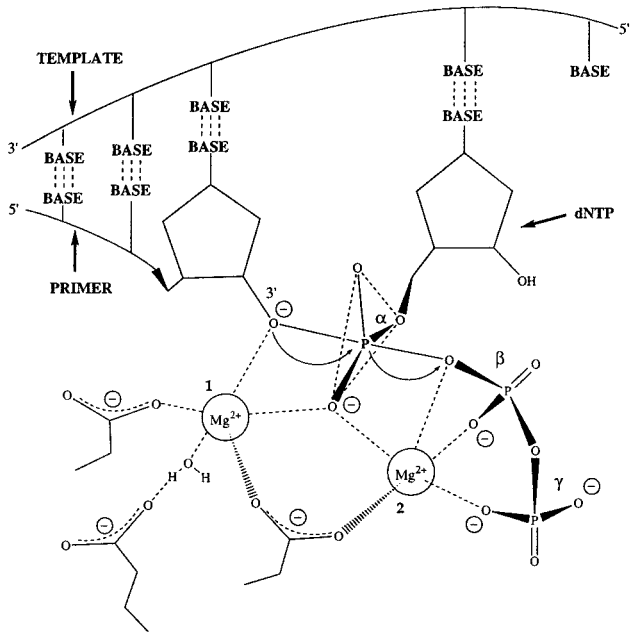


FIG. 5. Intermediate (or transition state) in the mechanism proposed for the polymerase reaction, involving catalysis mediated by two divalent metal ions. For a detailed description, see the text. The carboxylate ligands are in generic positions not intended to represent any particular polymerase. This figure (reprinted from *Science* [43] with permission from the publisher) is based on the original detailed proposal for a polymerase mechanism (42) but has been modified to reflect additional interactions of the dNTP β- and γ-phosphates with Mg^{2+} ion 2, as observed in Pol β complexes containing nucleotides (34, 40).

(the attacking nucleophile) in the Pol β complex are entirely consistent with the mechanism in Fig. 5.

dNTP binding interactions and their role in the polymerase reaction. Although no structural data exist for the corresponding ternary complex, mutagenesis experiments on Klenow fragment have provided some clues about the interactions that take place between the dNTP and active-site side chains as the polymerase reaction proceeds. Consistent with the expectation that the α-phosphate, the site of nucleophilic attack, must be bound at the catalytic center, kinetic experiments indicate that the α-phosphate is close to Asp-882 (37). Other portions of the molecule appear to contact Phe-762, Lys-758, and Arg-754 on the exposed face of helix O on the fingers subdomain (3). The contact with Arg-754, which probably involves the dNTP β- or γ-phosphate, affects the rate of chemical catalysis, perhaps by interacting with the PP_i leaving group as the phosphoryl transfer reaction proceeds. These assignments of dNTP contacts, though consistent with the experimental data, should still be considered tentative in the absence of a ternary complex structure. As we have pointed out elsewhere (23), very little of the data on dNTP binding to polymerases can be interpreted unambiguously. The kinetic studies of nucleotide utilization by mutant polymerases do not distinguish a direct effect of the altered side chain on a nucleotide contact from an indirect effect caused by a change in the way the unpaired template strand is presented for base pairing. (In our studies of Klenow fragment, mentioned above, we have assumed that the largest effects are most likely to result from loss of direct contacts to the dNTP.) Similarly, in HIV-1 reverse transcriptase, many of the mutations that influence the response to nucleoside analog inhibitors probably do so by repositioning the template strand. Thus, residues whose mutation makes the enzyme resistant to zidovudine (AZT) and ddI are either located in or interacting with a pair of antiparallel β-strands in the fingers subdomain that are proposed to interact with the unpaired template strand beyond the site of synthesis (26); recent biochemical studies of mutant proteins support this proposed interaction (10). Other structural and chemical cross-linking studies of dNTP binding to polymerases have focused on the binary enzyme-dNTP complex, which is not catalytically competent and therefore may not provide a complete picture of the interactions that take place in the enzyme-DNA-dNTP ternary complex. Nevertheless, most of the experiments designed to probe the interaction of polymerases with the incoming nucleotide have implicated residues on the fingers and the neighboring portion of the palm subdomain. Given that the fingers subdomain is thought to bind the template strand, a reasonable working hypothesis is that some residues from the fingers side of the cleft are involved indirectly in nucleotide binding by virtue of their role in orienting the template strand while others interact directly with the nucleotide.

NONHOMOLOGOUS POLYMERASE STRUCTURES

Insights provided by the Pol β structure. The structural data for Pol β that have been published recently (14, 34, 40) are particularly significant in that they include the first cocrystal structure of a polymerase ternary complex (34). The detailed picture of both substrates, DNA and nucleotide, at the polymerase active site has provided valuable evidence in favor of the two-metal-ion polymerase mechanism, which was originally proposed largely by analogy with the mechanism of the 3′-5′ exonuclease of Klenow fragment (5, 42). However, comparison of the entire Pol β structure with the three previously determined polymerase structures has not been straightforward, since there are two possible orientations (differing by a 180° rotation) in which the polymerase cleft of Pol β can be aligned with the other three structures, and there is disagreement as to the choice of orientation (43). Two features of the Pol β structure contribute to the difficulty in relating it to the structures of Klenow fragment, reverse transcriptase, and T7 RNA polymerase: (i) the polymerase domain of Pol β is clearly not homologous to that of the others and (ii) the molecule does not have additional functional domains analogous to those in the other polymerases which could provide clues as to the appropriate alignment of the four structures.

The palm subdomain of Pol β shows a superficial resemblance to the palm subdomains in the other polymerase structures, and this similarity guided initial attempts at aligning the structures (14, 40). However, the resulting alignment (shown schematically in Fig. 6) revealed some profound differences between Pol β and the other structures, indicating that the Pol β palm represents a new class of palm subdomain structures. Superposition of the two β-strands which bracket the motif C carboxylates, together with the adjoining two α-helices, places the trio of catalytic carboxylates in Pol β in a position similar to those in the other three structures; however, here the similarity ends. Whereas motif C in Klenow fragment and reverse transcriptase contains two adjacent carboxylates within a β-hairpin, Pol β has two carboxylates (whose importance has been demonstrated by mutagenesis [13]) separated by one residue on an extended β-strand. The observed correlation of sequence, -DD- versus -DxD-, with secondary structure seems necessary in order to place both carboxylate side chains on the same side of the β-sheet at the base of the polymerase cleft; thus, two adjacent side chains will be on the same side of the β-sheet if they are situated at a β-turn but not if they are present in an extended β-strand. Additional differences become apparent when the position of the third active-site car-

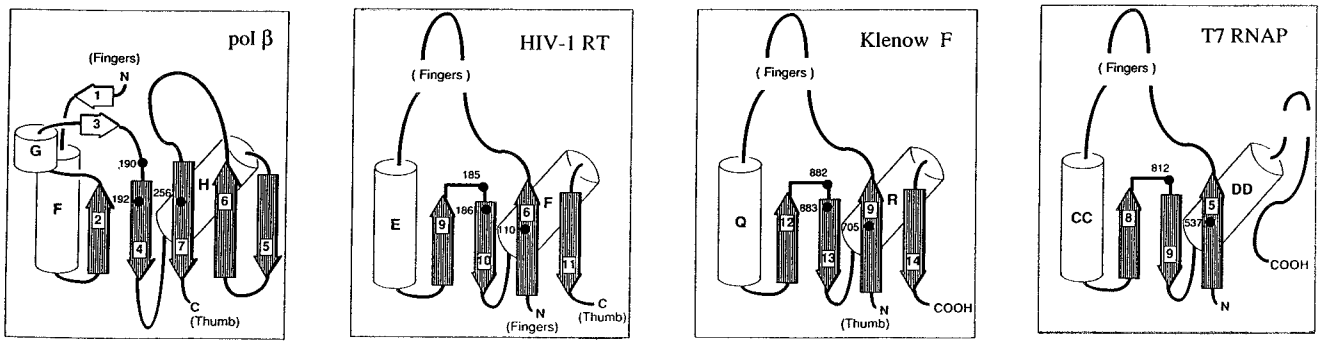


FIG. 6. Schematic representation illustrating the different connectivities in the palm subdomain of Pol β, HIV-1 reverse transcriptase (RT), Klenow fragment (F), and T7 RNA polymerase (RNAP) (figure kindly provided by Z. Hostomsky; figure reprinted from *Cell* [14] with permission of the publisher). The β-strands in the central β-sheet are depicted as numbered hatched arrows; helices are depicted as lettered cylinders. The positions of the carboxylate residues at the catalytic site are shown as black circles accompanied by residue numbers. In reverse transcriptase, Klenow fragment, and T7 RNA polymerase, the fingers loops at the top of the diagrams represent large departures of the main chain from the palm into the fingers subdomain.

boxylate of Pol β is examined. As described earlier, this residue is not the motif A carboxylate identified in the alignment of Delarue et al. (16) but is in fact located C terminal to motif C (14). Moreover, the β-strand containing this residue, though occupying a similar position to its counterpart in the other structures, runs in the opposite direction. In our view, these significant differences in the topology of the Pol β palm subdomain have several consequences. First, they argue strongly against an evolutionary relationship between Pol β and the other three polymerases; second, they negate any argument necessitating the alignment of Pol β with the other structures using the few remaining similar features within the palm subdomains. It has been pointed out that the one small structural motif, α-β-β-α, that is truly superimposable when Pol β is compared with the other structures is found in a number of other unrelated protein structures (40). Thus, this nonspecific feature appears insufficient on its own to mandate an alignment.

Examination of the Pol β ternary complex structure (34) also points out a major difficulty with aligning Pol β with Klenow fragment, reverse transcriptase, and T7 RNA polymerase in the orientation described above, since the direction in which the DNA template-primer approaches the polymerase active site in the Pol β ternary complex is opposite to that inferred from a variety of data (including two DNA cocrystal structures) for the other polymerases. Pelletier et al. (34) have argued on the basis of the Pol β data that the DNA binding models previously derived for the other three polymerases are incorrect. We and others continue to favor the earlier DNA binding models for the other polymerases, given the large body of supporting data (for recent reviews, see references 1 and 43). We have sought to reconcile the Pol β structural data by suggesting that this polymerase should be aligned with the other polymerases in the opposite orientation to that described above, i.e., by aligning the DNA substrates, rather than the nonhomologous enzyme structures (Fig. 7) (43). When Pol β is aligned in this way with reverse transcriptase (the only other polymerase for which appropriate DNA cocrystal data are available [21]), superposition of the two DNA primer termini produces a sufficiently close coincidence of the active-site carboxylates that the catalytic pair of metal ions can then be coordinated in the identical position relative to the primer terminus in both structures. (Note that the metal ions are not present in any of the reverse transcriptase crystal structures.) Thus, the identical geometrical arrangement of substrate molecules and catalytic groups (in this case the metal ions) can be achieved at the polymerase site even though the underlying protein scaffold and even the coordination geometries of the metal ions are quite distinct in the two structures. The idea of a similar active-site geometry being formed in two quite different structural contexts is in keeping with the idea that Pol β (which, as already described, is clearly not homologous to the other polymerases) must have achieved the same polymerase reaction mechanism by a process of convergent evolution.

In an alternative explanation, Arnold and coworkers have argued recently (1) that the alignment originally proposed for Pol β is the correct one but that the DNA oligonucleotide in the Pol β ternary complex may, for a variety of reasons, be bound to a binding site for duplex DNA downstream of the gap at which Pol β acts, resulting in an orientation of the DNA opposite to that appropriate for catalysis. We believe that this explanation is unlikely on the basis of the following considerations. (i) The position of the pair of metal ions relative to the dNTP phosphates effectively defines the direction of DNA synthesis as being that observed by Pelletier et al. (34). (ii) The dNTP α-phosphate must be bound adjacent to the 3′ hydroxyl of the primer strand; if the primer terminus in the cocrystal were in fact bound in the position that would normally be used by a DNA 5′ end, then one would not expect to see the dNTP in such close proximity. (iii) A DNA binding site will be complementary to or "fit" with only the correct polarity of the bound strand. A 3′ primer terminus would be unlikely to bind in the position appropriate for the 5′ phosphate of a strand having the opposite polarity.

Consideration of the way in which the Pol β structure relates to the other three polymerase structures has led to a greater understanding of what constitutes a polymerase domain. While the overall "palm-fingers-thumb" arrangement may well turn out to be a general feature of polymerase structures, the Pol β structure shows that we should not always expect to see as much similarity in the palm subdomain structures as was seen in the first three polymerase structures to be determined. The palm subdomain structure needs to generate an appropriate surface to accommodate the substrate molecules and to present the catalytic pair of metal ions in the appropriate geometrical arrangement relative to the primer 3′-hydroxyl and the dNTP α-phosphate, and it appears that there may be several quite distinct structures that meet these requirements. The positions of the carboxylates need not be identical (relative to the substrate molecules) for all polymerases, since the metal ion spacing appropriate for catalysis of phosphoryl transfer can be achieved with a variety of coordination patterns (as

FIG. 7. Backbone crystal structures of Klenow fragment, HIV-1 reverse transcriptase, T7 RNA polymerase, and DNA polymerase β (figure reprinted from *Science* [43] with permission of the publisher). The first three proteins were aligned by superposition of their homologous palm subdomains. Pol β was aligned by superimposing the terminal three phosphates of the bound primer strand (34) on the corresponding three phosphates of the primer strand bound to reverse transcriptase (21). The optimal alignment of the other polymerases on Pol β will be possible only when the positions of the two metal ions, the dNTP α-phosphate, and the 3′-OH of the primer strand are known in each case. The experimentally observed DNA backbones are shown as white coils for the primer strands and gray coils for the template strands. The Cα positions of the catalytic carboxylic acid residues are shown as yellow spheres. Functionally analogous interactions are seen between the primer-template and the fingers and thumb subdomains in all four polymerases. The template strand bound to reverse transcriptase has been extended by model building (short gray segments) to show that it interacts with the fingers subdomain, as is also the case for the template strand in the Pol β complex (using the changed naming of subdomains resulting from our alternative alignment).

can be seen, for example, by comparing the polymerase active site of Pol β with the 3′-5′ exonuclease active site of Klenow fragment [5, 34]). The functional analogy among the fingers and thumb subdomains in the four structures is also likely to be important, since these subdomains play an important role in binding and orienting the substrates, thus setting up the correct geometry in relation to the metal ions at the polymerase active site. In this respect, it is worth noting that the alignment we have proposed for the Pol β structure relative to those of the other polymerases (Fig. 7), though deemphasizing the apparent similarities in the palm subdomain, reinforces the functional analogies in the fingers and thumb subdomains (43).

Predictions for other polymerases. In light of the ideas discussed above, what predictions can be made about the structures of polymerases from families where no prototype structure is yet available? The available evidence suggests that the sequence alignment of Delarue et al. (16) provides a reasonable starting point for the identification of active-site residues. Thus, site-directed mutagenesis studies have demonstrated the importance of the carboxylates of motifs A and C in several members of the DNA polymerase α-family (6, 7, 12, 22, 24) and one RNA-dependent RNA polymerase (39). The Pol β structure hints at the variety of different palm subdomain structures that may be found in polymerases from nonhomologous families, though until more nonhomologous polymerase structures have been solved, we cannot predict how many different ways there may be of constructing a palm subdomain. Comparison between Pol β and the other three polymerase structures suggests at least two places where structural variability might be expected. One is the arrangement of the pair of motif C carboxylates, and here it has been suggested that the extended loop seen in the Pol β structure may be considered a prototype structure for other polymerases such as those of the Pol α family where the motif C carboxylates are also separated by one residue (14). The second source of differences between polymerase structures is likely to be the identity of the active-site carboxylate that is not derived from motif C and the connectivity of the secondary structure element on which this carboxylate is located. Although in Klenow fragment, reverse transcriptase, and T7 RNA polymerase, this other active-site carboxylate is derived from the conserved sequence motif A of Delarue et al. (16), in Pol β the alignment of Delarue et al. did not predict correctly the identity of the third active-site carboxylate. In the Pol α family, recent mutagenesis results raise the possibility that, in this family also, the invariant aspartate of motif A might not be one of the catalytically important car-

boxylates. Thus, although the D--SLYPS consensus sequence is strongly conserved in the large number of sequences of the Pol α family (11), mutagenesis of the aspartate residue has given mixed results, causing a drastic reduction in polymerase activity in ϕ29 DNA polymerase (7) but having very little effect in human DNA polymerase α (17). Clearly, structural information will be invaluable in interpreting these contradictory results.

CONCLUSIONS

The structural studies conducted in recent years have led to the conclusion that the polymerase reaction mechanism is likely to be the same (a phosphoryl transfer mediated by a pair of appropriately spaced divalent metal ions) throughout the polymerase superfamily. The polymerase domain serves to establish the correct geometrical arrangement of substrate molecules and metal ions at the active site so that catalysis can occur, but this purpose can be achieved in different structural ways. This appreciation of the structural diversity of this group of enzymes has been possible only through the comparison of the structures of several polymerases, including some that are not homologous to one another. Eventually, to understand the full repertoire of structural elements used by these enzymes, it will be necessary to have at least one structure determined from each polymerase family. In spite of this structural variability, the generality of the polymerase mechanism nevertheless means that, for some experimental purposes, simpler and experimentally more tractable polymerases can serve as valid models for more complex enzymes.

ACKNOWLEDGMENTS

We are grateful to the many coworkers who have shared our investigations of polymerase structure and function over the years. We also thank Nigel Grindley for comments on the manuscript, Joe Jäger for help in preparing some of the figures, and Kate Tatham for her patience in preparing the manuscript.

Polymerase research in our laboratories is supported by NIH grants GM-28550 and GM-39546, ACS grant BE-52G, and HHMI funding.

REFERENCES

1. **Arnold, E., J. Ding, S. H. Hughes, and Z. Hostomsky.** 1995. Structures of DNA and RNA polymerases and their interactions with nucleic acid substrates. Curr. Opin. Struct. Biol. **5:**27–38.
2. **Arnold, E., A. Jacobo-Molina, R. G. Nanni, R. L. Williams, X. Lu, J. Ding, A. D. Clark, Jr., A. Zhang, A. L. Ferris, P. Clark, A. Hizi, and S. H. Hughes.** 1992. Structure of HIV-1 reverse transcriptase/DNA complex at 7 Å resolution showing active site locations. Nature (London) **357:**85–89.
3. **Astatke, M., N. D. F. Grindley, and C. M. Joyce.** 1995. Deoxynucleoside triphosphate and pyrophosphate binding sites in the catalytically competent ternary complex for the polymerase reaction catalyzed by DNA polymerase I (Klenow fragment). J. Biol. Chem. **270:**1945–1954.
4. **Beese, L. S., V. Derbyshire, and T. A. Steitz.** 1993. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. Science **260:**352–355.
5. **Beese, L. S., and T. A. Steitz.** 1991. Structural basis for the 3′-5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. EMBO J. **10:**25–33.

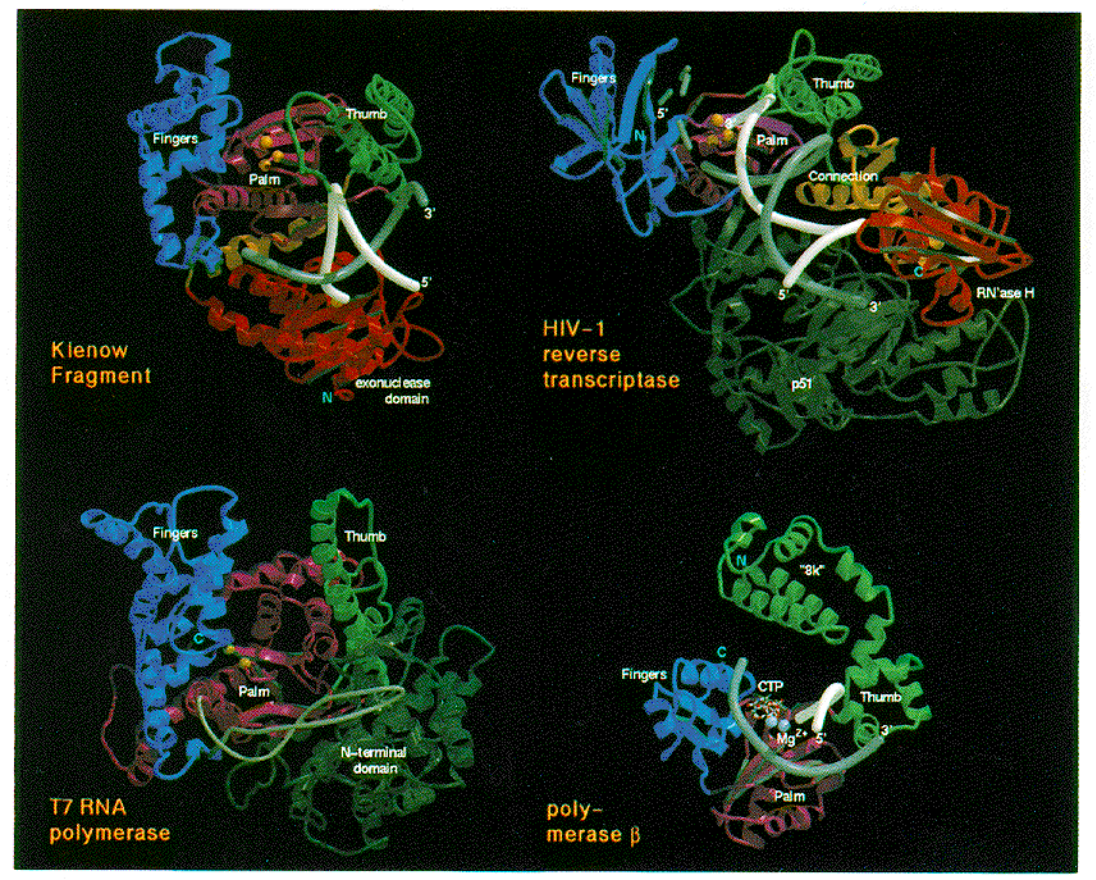
5a. **Beese, L. S., and T. A. Steitz.** Unpublished data.

6. **Bernad, A., J. M. Lázaro, M. Salas, and L. Blanco.** 1990. The highly conserved amino acid sequence motif Tyr-Gly-Asp-Thr-Asp-Ser in α-like DNA polymerases is required by phage ϕ29 DNA polymerase for protein-primed initiation and polymerization. Proc. Natl. Acad. Sci. USA **87:**4610–4614.
7. **Blasco, M. A., J. M. Lázaro, L. Blanco, and M. Salas.** 1993. ϕ29 DNA polymerase active site: residue Asp^{249} of conserved amino acid motif "Dx_2SLYP" is critical for synthetic activities. J. Biol. Chem. **268:**24106–24113.
8. **Bonner, G., D. Patra, E. M. Lafer, and R. Sousa.** 1992. Mutations in T7 RNA polymerase that support the proposal for a common polymerase active site structure. EMBO J. **11:**3767–3775.
9. **Boyer, P. L., A. L. Ferris, and S. H. Hughes.** 1992. Cassette mutagenesis of the reverse transcriptase of human immunodeficiency virus type 1. J. Virol. **66:**1031–1039.
10. **Boyer, P. L., C. Tantillo, A. Jacobo-Molina, R. G. Nanni, J. Ding, E. Arnold, and S. H. Hughes.** 1994. Sensitivity of wild-type human immunodeficiency virus type 1 reverse transcriptase to dideoxynucleotides depends on template length: the sensitivity of drug-resistant mutants does not. Proc. Natl. Acad. Sci. USA **91:**4882–4886.
11. **Braithwaite, D. K., and J. Ito.** 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res. **21:**787–802.
12. **Copeland, W. C., and T. S.-F. Wang.** 1993. Mutational analysis of the human DNA polymerase α: the most conserved region in α-like DNA polymerases is involved in metal-specific catalysis. J. Biol. Chem. **268:**11028–11040.
13. **Date, T., S. Yamamoto, K. Tanihara, Y. Nishimoto, and A. Matsukage.** 1991. Aspartic acid residues at positions 190 and 192 of rat DNA polymerase β are involved in primer binding. Biochemistry **30:**5286–5292.
14. **Davies, J. F., R. J. Almassy, Z. Hostomska, R. A. Ferre, and Z. Hostomsky.** 1994. 2.3 Å crystal structure of the catalytic domain of DNA polymerase β. Cell **76:**1123–1133.
15. **Davies, J. F., Z. Hostomska, Z. Hostomsky, S. R. Jordan, and D. A. Matthews.** 1991. Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. Science **252:**88–95.
16. **Delarue, M., O. Poch, N. Tordo, D. Moras, and P. Argos.** 1990. An attempt to unify the structure of polymerases. Protein Eng. **3:**461–467.
17. **Dong, Q., W. C. Copeland, and T. S.-F. Wang.** 1993. Mutational studies of human DNA polymerase α: identification of residues critical for deoxynucleotide binding and misinsertion fidelity of DNA synthesis. J. Biol. Chem. **268:**24163–24174.
18. **Freemont, P. S., J. M. Friedman, L. S. Beese, M. R. Sanderson, and T. A. Steitz.** 1988. Cocrystal structure of an editing complex of Klenow fragment with DNA. Proc. Natl. Acad. Sci. USA **85:**8924–8928.
19. **Heringa, J., and P. Argos.** 1994. Evolution of viruses as recorded by their polymerase sequences. *In* S. S. Morse (ed.), Evolutionary biology of viruses, p. 87–103. Raven Press, New York.
20. **Ito, J., and D. K. Braithwaite.** 1991. Compilation and alignment of DNA polymerase sequences. Nucleic Acids Res. **19:**4045–4057.
21. **Jacobo-Molina, A., J. Ding, R. G. Nanni, A. D. Clark, Jr., X. Lu, C. Tantillo, R. L. Williams, G. Kamer, A. L. Ferris, P. Clark, A. Hizi, S. H. Hughes, and E. Arnold.** 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. Proc. Natl. Acad. Sci. USA **90:**6320–6324.
22. **Joung, I., M. S. Horwitz, and J. A. Engler.** 1991. Mutagenesis of conserved region I in the DNA polymerase from human adenovirus serotype 2. Virology **184:**235–241.
23. **Joyce, C. M., and T. A. Steitz.** 1994. Function and structure relationships in DNA polymerases. Annu. Rev. Biochem. **63:**777–822.
24. **Jung, G., M. C. Leavitt, M. Schultz, and J. Ito.** 1990. Site-specific mutagenesis of PRD1 DNA polymerase: mutations in highly conserved regions of the family B DNA polymerase. Biochem. Biophys. Res. Commun. **170:**1294–1300.
25. **Kim, E. E., and H. W. Wyckoff.** 1991. Reaction mechanism of alkaline phosphatase based on crystal structures. J. Mol. Biol. **218:**449–464.
26. **Kohlstaedt, L. A., J. Wang, J. M. Friedman, P. A. Rice, and T. A. Steitz.** 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science **256:**1783–1790.
27. **Kumar, A., J. Abbotts, E. M. Karawya, and S. H. Wilson.** 1990. Identification and properties of the catalytic domain of mammalian DNA polymerase β. Biochemistry **29:**7156–7159.
28. **Kumar, A., S. G. Widen, K. R. Williams, P. Kedar, R. L. Karpel, and S. H. Wilson.** 1990. Studies of the domain structure of mammalian DNA polymerase β: identification of a discrete template binding domain. J. Biol. Chem. **265:**2124–2131.
29. **Larder, B. A., S. D. Kemp, and D. J. M. Purifoy.** 1989. Infectious potential of human immunodeficiency virus type 1 reverse transcriptase mutants with altered inhibitor sensitivity. Proc. Natl. Acad. Sci. USA **86:**4803–4807.
30. **Larder, B. A., D. J. M. Purifoy, K. L. Powell, and G. Darby.** 1987. Site-specific mutagenesis of AIDS virus reverse transcriptase. Nature (London) **327:**716–717.

30a. **Matsumoto, Y., and K. Kim.** 1995. Excision of deoxyribose phosphate residues by DNA polymerase β during DNA repair. Science **269:**699–702.

31. **McAllister, W. T., and C. A. Raskin.** 1993. The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. Mol. Microbiol. **10:**1–6.
32. **Ollis, D. L., P. Brick, R. Hamlin, N. G. Xuong, and T. A. Steitz.** 1985. Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. Nature (London) **313:**762–766.
33. **Osumi-Davis, P. A., M. C. de Aguilera, R. W. Woody, and A.-Y. M. Woody.** 1992. Asp537, Asp812 are essential and Lys631, His811 are catalytically significant in bacteriophage T7 RNA polymerase activity. J. Mol. Biol. **226:**37–45.
34. **Pelletier, H., M. R. Sawaya, A. Kumar, S. H. Wilson, and J. Kraut.** 1994. Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer, and ddCTP. Science **264:**1891–1903.
35. **Piccirilli, J. A., J. S. Vyle, M. H. Caruthers, and T. R. Cech.** 1993. Metal ion catalysis in the *Tetrahymena* ribozyme reaction. Nature (London) **361:**85–88.

36. **Poch, O., I. Sauvaget, M. Delarue, and N. Tordo.** 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. **8:**3867–3874.
37. **Polesky, A. H., M. E. Dahlberg, S. J. Benkovic, N. D. F. Grindley, and C. M. Joyce.** 1992. Side chains involved in catalysis of the polymerase reaction of DNA polymerase I from *Escherichia coli*. J. Biol. Chem. **267:**8417–8428.
38. **Polesky, A. H., T. A. Steitz, N. D. F. Grindley, and C. M. Joyce.** 1990. Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from *Escherichia coli*. J. Biol. Chem. **265:** 14579–14591.
39. **Sankar, S., and A. G. Porter.** 1992. Point mutations which drastically affect the polymerization activity of encephalomyocarditis virus RNA-dependent RNA polymerase correspond to the active site of *Escherichia coli* DNA polymerase I. J. Biol. Chem. **267:**10168–10176.
40. **Sawaya, M. R., H. Pelletier, A. Kumar, S. H. Wilson, and J. Kraut.** 1994. Crystal structure of rat DNA polymerase β: evidence for a common polymerase mechanism. Science **264:**1930–1935.
41. **Sousa, R., Y. J. Chung, J. P. Rose, and B.-C. Wang.** 1993. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. Nature (London) **364:**593–599.
42. **Steitz, T. A.** 1993. DNA- and RNA-dependent DNA polymerases. Curr. Opin. Struct. Biol. **3:**31–38.
43. **Steitz, T. A., S. Smerdon, J. Jäger, and C. M. Joyce.** 1994. A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. Science **266:**2022–2025.
44. **Steitz, T. A., S. Smerdon, J. Jäger, J. Wang, L. A. Kohlstaedt, J. M. Friedman, L. S. Beese, and P. A. Rice.** 1993. Two polymerases: HIV reverse transcriptase and the Klenow fragment of *E. coli* DNA polymerase I. Cold Spring Harbor Symp. Quant. Biol. **58:**459–504.
45. **Steitz, T. A., and J. A. Steitz.** 1993. A general two-metal-ion mechanism for catalytic RNA. Proc. Natl. Acad. Sci. USA **90:**6498–6502.
46. **Wong, S. W., A. F. Wahl, P.-M. Yuan, N. Arai, B. E. Pearson, K.-I. Arai, D. Korn, M. W. Hunkapiller, and T. S.-F. Wang.** 1988. Human DNA polymerase α gene expression is cell proliferation dependent and its primary structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. EMBO J. **7:**37–47.
47. **Yang, W., W. A. Hendrickson, R. J. Crouch, and Y. Satow.** 1990. Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. Science **249:**1398–1405.


Fingers
Thumb
Palm
3'
5'
N
exonuclease domain
Klenow Fragment
Fingers
Thumb
N
5'
Palm
Connection
C
RN'ase H
5'
3'
p51
HIV-1 reverse transcriptase
Fingers
Thumb
C
Palm
N-terminal domain
T7 RNA polymerase
N
"8k"
C
Fingers
CTP
Thumb
Mg^{2+}
5'
3'
Palm
poly-merase β