

tion state (22). On the enzymatic side of the same transition state, the catalytic groups participating in proton abstraction are stabilized by the LBHB.

In the LBHB-facilitated mechanism (Fig. 2), the catalytic advantage conferred by the LBHB should be lost by its disruption. Accordingly, *N*-methylation of His⁵⁷ in α -chymotrypsin decreases its acylation rate by 2×10^5 (23). Moreover, the overall catalytic activity of the trypsin mutant Asp¹⁰² \rightarrow Asn (chymotrypsin numbering) was reduced to one ten-thousandth of its original value (24). These experiments support but do not prove the LBHB-facilitated general base mechanism.

LBHBs have been postulated to participate in the enzymatic enolization of aldehydes, ketones, and esters (25–27). Spectroscopic evidence for LBHBs in these enzymes has so far not been reported. It is important to verify LBHBs spectroscopically because electrostatic stabilization can also facilitate the enolization of ketones and esters (28, 29). The assignment of LBHBs in diverse enzymatic processes should be supported by spectroscopic information such as that obtained for serine proteases.

REFERENCES AND NOTES

- F. Hibbert and J. Emsley, *Adv. Phys. Org. Chem.* **26**, 255 (1990).
- J. Emsley, *Chem. Soc. Rev.* **9**, 91 (1980).
- T. Zeegers-Huyskens and L. Sobczyk, *J. Mol. Liq.* **46**, 263 (1990).
- Z. Dega-Szafran and E. Dulewicz, *Org. Magn. Reson.* **16**, 214 (1981).
- G. Gunnarsson, H. Wennerström, W. Egan, S. Forsén, *Chem. Phys. Lett.* **38**, 96 (1976).
- B. Brezinski and M. Szafran, *Org. Magn. Reson.* **15**, 78 (1981).
- Photoisomerization was carried out according to S. Edelbacher and Fr. Heitz [*Hoppe-Seyler's Z. Physiol. Chem.* **279**, 63 (1943)]. Chromatography was as described [M. Morrison and D. Avnir, *J. Chromatogr.* **183**, 83 (1980)] [¹H NMR in DMSO-*d*₆ (δ , multiplicity, integration, coupling constant): 5.68 ppm, doublet, 1 H, 12.3 Hz; 6.90 ppm, doublet, 1 H, 12.6 Hz; 7.67 ppm, singlet, 1 H; 8.15 ppm, singlet, 1 H; 12.9 ppm, singlet, 1 H; 17.4 ppm, singlet, 1 H].
- G. Robillard and R. G. Shulman, *J. Mol. Biol.* **71**, 507 (1972).
- J. H. Clark, M. Green, R. G. Madden, *J. Chem. Soc. Chem. Commun.* **1983**, 136 (1983).
- J. L. Markley, *Biochemistry* **17**, 46 (1978).
- W. W. Bachovchin, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7948 (1985).
- A. J. N. Murthy, S. N. Bhat, C. N. R. Rao, *J. Chem. Soc. A* **1970**, 1251 (1970).
- D. M. Blow, J. J. Birkoft, B. S. Hartley, *Nature* **221**, 337 (1969).
- G. A. Rogers and T. C. Bruice, *J. Am. Chem. Soc.* **96**, 2473 (1974).
- S. C. Zimmerman, J. S. Korthals, K. D. Cramer, *Tetrahedron* **47**, 2649 (1991).
- J. L. Markley and I. B. Ibañez, *Biochemistry* **17**, 4627 (1978).
- A. A. Kossiakoff and S. A. Spencer, *ibid.* **20**, 6462 (1981).
- F. C. Kokesh and F. H. Westheimer, *J. Am. Chem. Soc.* **93**, 7270 (1971).
- G. Robillard and R. G. Shulman, *J. Mol. Biol.* **86**, 541 (1974).
- T.-C. Liang and R. H. Abeles, *Biochemistry* **26**, 7603 (1987).
- J. D. Robertus, J. Kraut, R. A. Alden, J. J. Birkoft, *ibid.* **11**, 4293 (1972).
- The oxyanion of the tetrahedral intermediate was recently postulated to be stabilized by LBHBs from main chain N–H bonds of both Gly¹⁹³ and Ser¹⁹⁵ [J. A. Gerlt and P. G. Gassman, *Biochemistry* **32**, 11943 (1993)]. This is questionable on three grounds: (i) One heteroatom participating simultaneously in two LBHBs is unprecedented. (ii) The p*K*_a of the OH group in the tetrahedral intermediate was taken to be 13.4, matching that of the peptide N–H bonds. Considering p*K*_a's for related molecules, 13.4 seems too high. The p*K*_a for ethanol is 16 and that for acetaldehyde hydrate is 13.5 (30), so that replacing one H on C-1 of ethanol with OH lowers the p*K*_a by 2.5 units. In the tetrahedral intermediate, there is a third electronegative atom bonded to the alcoholic carbon, which would further lower the p*K*_a to 11 or 11.5 and weaken the hydrogen bond. Any medium effect that perturbs this p*K*_a should have parallel effects on the p*K*_a's of the amide hydrogens, which are also neutral acids, and would not lead to matched p*K*_a's. (iii) NMR data on the tetrahedral adduct of *N*-acetyl-L-Leu-L-Phe trifluoromethyl ketone with Ser¹⁹⁵ of chymotrypsin indicates that its p*K*_a is approximately 4.9 (20).
- R. Henderson, *Biochem. J.* **124**, 13 (1971).
- C. S. Craik, S. Rocznik, C. Largman, W. J. Rutter, *Science* **237**, 909 (1987); S. Sprang *et al.*, *ibid.*, p. 905. Craik *et al.*, postulated that in the triad of the mutant Asp¹⁰² \rightarrow Asn (chymotrypsin numbering system), histidine exists in its tautomeric structure in which the imidazole proton is bonded to Ne2 and N δ 1 is hydrogen-bonded to the β -carboxamide nitrogen of Asn¹⁰². Tautomerization would abolish the general base reactivity of His⁵⁷ and loss of the LBHB would have a similar effect.
- W. W. Cleland, *Biochemistry* **31**, 317 (1992).
- J. A. Gerlt and P. G. Gassman, *J. Am. Chem. Soc.* **115**, 11552 (1993).
- W. W. Cleland and M. M. Kreevoy, *Science* **264**, 1887 (1994).
- J. P. Guthrie and R. Kluger, *J. Am. Chem. Soc.* **115**, 11569 (1993).
- C. J. Halkides, P. A. Frey, J. B. Tobin, *ibid.*, p. 3332.
- W. P. Jencks and J. Regenstein, in *CRC Handbook of Biochemistry and Molecular Biology*, H. A. Sober, Ed. (Chemical Rubber Co., Cleveland, OH, 1970), p. J187.
- We acknowledge valuable discussions with W. W. Cleland about LBHBs in enzymatic reactions. We thank G. B. Schuster and M. Zhang of the University of Illinois for hospitality and assistance in carrying out the photoisomerization of *trans*-urocanic acid in their laboratory. Supported by grant DK 28607 from the National Institute of Diabetes and Digestive and Kidney Diseases. This study made use of the National Magnetic Resonance Facility at Madison, which is supported by NIH grant RR02301 from the Biomedical Research Technology Program, National Center for Research Resources. Equipment in the facility was purchased with funds from the University of Wisconsin, the NSF Biological Instrumentation Program (grant DMB-8415048), NIH Biomedical Research Technology Program (grant RR02301), NIH Shared Instrumentation Program (grant RR02781), and the U.S. Department of Agriculture.

1 November 1993; accepted 25 April 1994

Crystal Structure of Rat DNA Polymerase β : Evidence for a Common Polymerase Mechanism

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

Structures of the 31-kilodalton catalytic domain of rat DNA polymerase β (pol β) and the whole 39-kilodalton enzyme were determined at 2.3 and 3.6 angstrom resolution, respectively. The 31-kilodalton domain is composed of fingers, palm, and thumb subdomains arranged to form a DNA binding channel reminiscent of the polymerase domains of the Klenow fragment of *Escherichia coli* DNA polymerase I, HIV-1 reverse transcriptase, and bacteriophage T7 RNA polymerase. The amino-terminal 8-kilodalton domain is attached to the fingers subdomain by a flexible hinge. The two invariant aspartates found in all polymerase sequences and implicated in catalytic activity have the same geometric arrangement within structurally similar but topologically distinct palms, indicating that the polymerases have maintained, or possibly re-evolved, a common nucleotidyl transfer mechanism. The location of Mn²⁺ and deoxyadenosine triphosphate in pol β confirms the role of the invariant aspartates in metal ion and deoxynucleoside triphosphate binding.

DNA polymerase β (pol β) is one of the four recognized DNA-directed DNA polymerases of the eukaryotic nucleus: α , β , δ , and ϵ . It has been studied primarily in vertebrates but homologs have also recently been discovered in lower eukaryotes such as yeast (1) and trypanosomes (2). Although

its full physiological role has not yet been established, several lines of evidence suggest that it is an essential housekeeping enzyme (3): the pol β gene is highly conserved and constitutively expressed; its activity has been correlated with repair of damaged DNA; and it alone has the ability to fill single-stranded DNA gaps smaller than six nucleotides (4).

As a single 335-residue polypeptide chain with no associated exonuclease or proofreading activity, pol β is a relatively uncomplicated target for mechanistic studies of the nucleotidyl transfer reaction—the coupling

M. R. Sawaya, H. Pelletier, J. Kraut, Department of Chemistry, University of California, San Diego, CA 92093-0317, USA.

A. Kumar and S. H. Wilson, Sealy Center for Molecular Science, University of Texas Medical Branch, Galveston, TX 77555-1051, USA.

of a nucleoside 5'-triphosphate (rNTP or dNTP) and the free 3'-OH of a primer strand to extend the primer by 1 nt and eliminate one molecule of inorganic pyrophosphate. Although this reaction is the principal unifying theme among the four categories of polymerases (DNA-directed DNA and RNA polymerases and also RNA-directed DNA and RNA polymerases), attempts to show this unification mechanistically have been thwarted by differences in the polymerases' biological roles, size, fidelity, processivity, drug sensitivity, and association with auxiliary domains or subunits. The conserved geometric arrangement of catalytic aspartic acid residues found in pol β and the three other structurally determined polymerases now provides compelling evidence that all the polymerases do indeed share a common mechanism of nucleotidyl transfer. Pelletier *et al.* have proposed a catalytic mechanism on the basis of the structure of pol β complexed with template-primer and ddCTP (5).

Rat pol β was expressed and purified from *Escherichia coli* as described (6). Crystals of the 31-kD domain (residues 88–335) were grown at room temperature from sitting drops prepared by mixing 5 μ l of protein solution with 5 μ l of buffer [7.5% (w/v) PEG 3350, 100 mM tris, pH 7.5, 15 mM SrCl₂]. The protein solution contained 39-kD pol β (15 mg/ml) dissolved in 100 mM Hepes, pH 7.5, and 10 mM (NH₄)₂SO₄. The reservoir contained 25% PEG 3350, 100 mM Hepes, pH 7.5, 15

mM SrCl₂. Crystal growth under these conditions required the presence of an infecting microbial protease that cleaved the polypeptide chain at the hinge connecting the 8-kD

and 31-kD domains (7). Crystals of the 31-kD domain of pol β belong to space group *P*2₁2₁2 (*a* = 120.6, *b* = 63.5, and *c* = 38.2 Å) with one molecule per asymmetric unit. The best

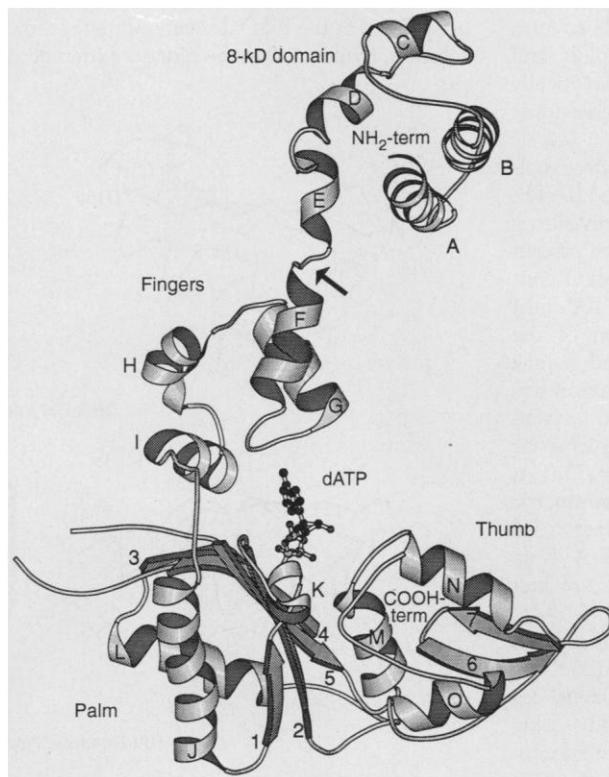


Fig. 1. Ribbon drawing (57) of rat DNA polymerase β . The individual domains-subdomains are composed of the following residues: 8-kD domain (1–87), fingers (88–151), palm (152–262), thumb (263–335). The arrow marks the location of the proteolytically sensitive hinge between the 31-kD catalytic domain and the 8-kD template binding domain. The ribbon diagram of KF and HIV RT are shown in the same orientation in (14). T7 RNAP is shown in (16). All four polymerases contain DNA binding channels composed of fingers, palm, and thumb regions. The 8-kD domain has no counterpart in KF, RT, or RNAP.

Table 1. Structure solution statistics for rat pol β . Intensity data were collected on a Xuong-Hamlin multiwire area detector system (48) and processed with the UCSF software package (49). The structure of the 31-kD domain was solved by multiple isomorphous replacement and anomalous scattering techniques (MIRAS). MIRAS phases were calculated to 2.7 Å resolution from two derivatives, K₂PtCl₄ and *p*-chloromercuriphenylsulfonic acid (PCMPS) (50), with overall isomorphous phasing powers of 2.8 and 1.7, respectively, and an overall anomalous phasing power of 0.7 and 0.6, respectively. Least squares phase refinement with combined heavy atom data yielded a figure of merit of 0.68 and solvent flattening (51) further improved the figure of merit to 0.74. A model containing 80% of the main chain atoms and 50% of the side chain atoms was then built into the resulting map. Following phase combination with program SIGMA A (52) the remaining main chain and side chains were easily traced with the TOM version of FRODO (53) installed on a Silicon Graphics workstation. TNT least squares positional and B-factor refinement (54) reduced the *R* factor for the 31-kD domain

from 45% to 19.3% for all reflections between 20 Å and 2.3 Å resolution. The current model includes 90 water molecules. Refinement statistics and a Ramachandran plot (not shown) indicate that the model has good geometry; only Glu³⁰⁹, part of a disordered loop, falls in a disallowed region. There is one cis-peptide bond, between Gly²⁷⁴ and Ser²⁷⁵. The structure of the full 39-kD apoenzyme was solved by application of the MERLOT molecular replacement program with the 31-kD domain as a search model (55). X-PLOR (56) rigid body refinement and TNT positional refinement lowered the *R* factor from 47% to 25.5%. The model for the 8-kD domain was taken from the *P*6₁ ternary complex (5) and manually placed in positive *F*_o - *F*_c α_c electron density, then the entire structure was subjected to X-PLOR rigid body refinement and TNT atomic positional refinement, lowering the *R* factor to 19.3%. No water molecules have as yet been included in the structure. An *F*_{o(dATP)} - *F*_{o(apo)} difference Fourier map revealed the location of dATP as a 6 σ positive electron density peak. The dATP was then modeled and the structure refined to a current *R* factor of 19.9%.

Crystal	Size (kD)	Space group	<i>d</i> _{min} (Å)	<i>I</i> / σ^*	Data collection				Refinement			
					Reflections		Completeness		<i>R</i> _{sym} † (%)	rms deviation		<i>R</i> ‡ factor (%)
					Total	Unique	Isomorphous (%)	Anomalous (%)		Bond (Å)	Angle (deg)	
Native	31	<i>P</i> 2 ₁ 2 ₁ 2	2.3	1.9	33939	13634	93.1		4.5	0.021	3.0	19.3
K ₂ PtCl ₄	31	<i>P</i> 2 ₁ 2 ₁ 2	2.7	3.1	36598	7901	92.3	84.0	6.1			
PCMPS	31	<i>P</i> 2 ₁ 2 ₁ 2	2.7	1.8	29933	7825	91.4	81.2	6.9			
Mn ²⁺	31	<i>P</i> 2 ₁ 2 ₁ 2	2.8	1.8	22234	6929	94.8	86.6	6.7	0.020	3.0	19.0
Native	39	<i>P</i> 2 ₁	3.6	1.8	14621	4153	99.5		7.2	0.017	2.9	19.3
dATP	39	<i>P</i> 2 ₁	3.9	1.8	12400	4069	91.0		6.7	0.019	3.0	19.9

*Average ratio of observed intensity to σ for the highest resolution shell of reflections. 20 Å and *d*_{min}.

†*R*_{sym} = $\sum |I_{obs} - I_{avg}| / \sum I_{avg}$.

‡Calculated with the use of reflections between

crystals grew to 0.9 by 0.4 by 0.4 mm and diffracted to a Bragg resolution of 2.3 Å. Isomorphous crystals containing Mn^{2+} were produced by soaking the apo 31-kD $P2_12_12$ crystals in 6 mM $MnCl_2$ for 3 days. After the structure was solved, we found that cloned and purified 31-kD domain was able to form the same $P2_12_12$ crystals more rapidly and reproducibly compared to the proteolytically cleaned enzyme, but required a higher ionic strength (8).

Crystals of the full 39-kD polymerase molecule, both with and without bound dATP, were grown at room temperature in sitting drops prepared by mixing 5 μ l of protein solution with 5 μ l of reservoir solution containing 7.5% (w/v) PEG 3350, 100 mM Hepes, pH 7.0. For crystallization of the dATP complex, 5 mM $MgCl_2$ and 5 mM dATP were added. The protein solution was the same as for the $P2_12_12$ crystals. Crystals grew within 2 weeks. They belong to space group $P2_1$ ($a = 39.2$, $b = 68.2$, $c = 75.0$ Å; $\beta = 91.8^\circ$) with one molecule per asymmetric unit. Crystals of the apoenzyme diffract to 3.6 Å and of the dATP complex to 3.9 Å. Data collection and refinement statistics are presented in Table 1.

Rat pol β comprises two domains, an NH_2 -terminal 8-kD domain (residues 1–87) and a $COOH$ -terminal 31-kD catalytic domain (residues 88–335) (Fig. 1). The 8-kD domain has been shown by affinity and photocrosslinking studies to bind single-stranded DNA, and consequently is thought to participate in template binding (9, 10). There are only two other polymerases containing sequences similar to that of the 8-kD domain; these are the terminal deoxynucleotidyltransferase (11) and polyadenylate polymerase (12). The 31-kD domain takes the form of a DNA binding channel similar in size and shape to those of the three polymerase structures determined so far, namely, the Klenow fragment (KF) of *E. coli* DNA polymerase I (13), HIV-1 reverse transcriptase (RT) (14, 15) and bacteriophage T7 RNA polymerase (RNAP) (16). Its dimensions are compatible with binding to B-form DNA. In KF, RT, RNAP, and now pol β , three subdomains make up the floor and sides of the DNA binding channel. They have been termed the fingers, palm, and thumb subdomains (Fig. 1) because models of polymerase-DNA complexes resemble a right hand holding a rod (13, 14). The 8-kD domain of pol β is attached to the 31-kD domain at the tip of the fingers subdomain. The 8-kD domain alone has no DNA polymerase activity while the 31-kD domain by itself has approximately 5% of the full activity. Evidently full activity requires that the two domains be covalently linked (6).

The 8-kD domain is composed of two pairs of antiparallel helices crossed at about 80° with respect to each other (Fig. 1) (17). The

hydrophilic loop or hinge connecting the 8-kD domain to the fingers subdomain extends away from the palm, leaving the channel open. However, the position of the 8-kD domain observed in the 39-kD structure must be but one of many possible nonproductive positions that the 8-kD domain may adopt in solution. Evidence that the hinge is extremely

sensitive to proteases except when template DNA is bound (6) suggests that the 8-kD domain may fold over the bound template. In fact, this domain moves closer to the palm in the complexes with template-primer (5) although perhaps primarily as a result of crystal packing forces rather than specific interactions with template-primer.

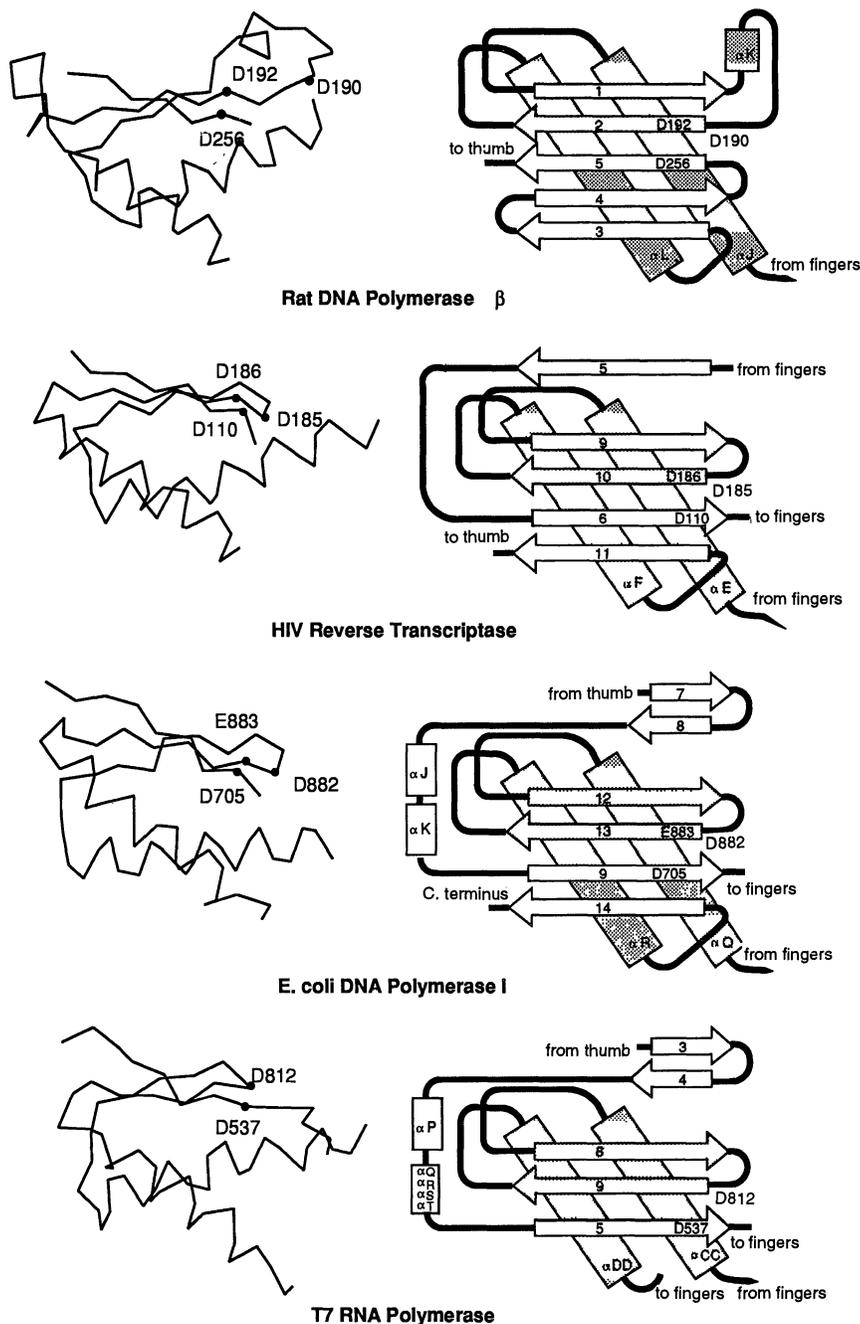


Fig. 2. (Left column) α -Carbon trace of structural elements common to all four polymerase palms: pol β , RT, KF, and RNAP. Structures were superposed then translated for viewing. The pseudo twofold axis of symmetry passes vertically through the center of the palm as shown here (21). (Right column) Topology diagrams of the four polymerases, pol β , RT, KF, and RNAP. The conserved acidic residues of motif A and C are labeled. Motifs A and C lie on antiparallel β strands in RT, KF, and RNAP, but they are parallel in pol β . The aspartic acids of motif A are Asp²⁵⁶ (pol β); Asp¹¹⁰ (RT); Asp⁷⁰⁵ (KF); and Asp⁵³⁷ (RNAP). The conserved residues of motif C are Asp¹⁹⁰, Asp¹⁸⁵ (pol β); Asp¹⁸⁵, Asp¹⁸⁶ (RT); Asp⁸⁸², Glu⁸⁸³ (KF); and Asp⁸¹² (RNAP). The topologically conserved region among the four polymerases is colored gray. This conserved region contains motif C.

The DNA binding channels of KF, RT, and RNAP resemble each other more closely than they resemble the binding channel of pol β . Indeed, the degree of structural similarity among KF, RT, and RNAP has led to their designation as "two sisters and their cousin" (18). In particular, the thumbs of KF, RT, and RNAP contain long (30 Å) antiparallel α helical structures which have been proposed to increase processivity by wrapping around the template-primer (15, 19). In contrast, the thumb of pol β is a small α + β structure, overlapping only the base of the other thumbs and part of the palms as defined by helices J and K in KF, β strands 12, 13, and 14 of RT, and helix P in RNAP. [In the accompanying paper, these overlapping structural elements are proposed to perform a common function, nucleotide discrimination (5)]. A similar disparity in structural design is found among the fingers subdomains. The channel walls formed by the fingers of KF, RT, and RNAP, though structurally divergent, have approximately the same gross dimensions, but the same wall in pol β is much longer and more flexible as it is formed from both the fingers and the 8-kD domain.

The remaining subdomain of the DNA binding channel, the palm, is the most structurally conserved (Fig. 2). The palm is formed from two α helices packed against one face of a β sheet. Such a fold is common in many proteins (see below) and has been termed a two layered α - β sandwich (20). The pol β palm superposes on KF, RT, and RNAP palms with root-mean-square (rms) deviations of 2.7, 2.5 and 2.6 Å, respectively, for 60 α -carbons (21). But, here too, pol

β differs from the other polymerases. The β sheets composing the palms of KF, RT, and RNAP are all antiparallel, but pol β contains an additional inserted parallel strand (Fig. 2). Because one of the conserved aspartic acid residues lies on this parallel strand (as is shown below), the topological difference in palms has implications for both the evolution of polymerases and the generality of a catalytic mechanism for nucleotidyl transfer.

Current arguments for a common mechanism of nucleotidyl transfer rest on three lines of evidence. (i) There are two universal sequence motifs among all four categories of polymerases, termed motifs A and C (22). Both motifs contain an invariant aspartic acid residue. In most polymerases, the aspartic acid of motif C is followed by another conserved Asp or Glu. These conserved residues have been identified for KF, RT, RNAP, and pol β (Fig. 2). (ii) Mutagenesis studies have demonstrated that mutations in these acid residues in KF, RT, and RNAP decrease k_{cat} of nucleotidyl transfer by 100 to 50,000 times (23–26). Similarly, conservative mutation of Asp¹⁹⁰ (motif C) to glutamic acid in pol β has also been shown to drop catalytic activity to 0.1% (27). Because the conserved residues have carboxylic acid side chains and the catalytic reaction is known to involve divalent metal ions these residues were thought to be coordinated by one or more metal ions. This hypothesis is confirmed by the structure of the pol β -Mn²⁺ complex (28). (iii) Structural studies further showed that the aspartic acids of motifs A and C in KF, RT, and RNAP are spatially clustered together in the palm so that they may be cooperatively coord-

inated by metal ions. Motifs A and C each consists of a single β strand, and the strands are antiparallel but separated in sequence by the fingers (Figs. 2 and 3). The invariant aspartic acid of motif C participates in a tight β turn at the NH₂-terminal end of its β strand, while the invariant aspartic acid of motif A lies at the carboxy end of its β strand, but spatially near the aspartic acid of motif C. The structure of pol β shows a similar spatial arrangement of carboxylic acid groups, confirming that these polymerases all contain a common active site geometry and therefore are likely to operate by the same catalytic mechanism. Evidence that this argument may include all other polymerases derives from a closer examination of the structural differences in the palms.

The palm of pol β differs from those of other polymerases in several important respects. (i) The two β strands of the catalytic site are parallel, not antiparallel as in KF, RT, and RNAP (Fig. 2). (ii) The two strands are not separated in sequence by some or all of the fingers subdomain. In fact, pol β is the only polymerase with sequentially continuous subdomains (Fig. 3). (iii) Motif A follows motif C in sequence instead of preceding it. For this reason, although Delarue *et al.* (22) correctly identified Asp¹⁹⁰ and Asp¹⁹² as belonging to motif C, they could not identify motif A correctly; unexpectedly Asp²⁵⁶, not Asp¹⁷, is structurally equivalent to the conserved aspartic acid residues of motif A in KF, RT, and RNAP. (iv) Motif C does not participate in a β turn, but instead maintains the main chain torsion angles of a β sheet. That is why the two carboxylic acid side chains of motif C appear to be similarly positioned on the open face of the β sheet even though they are sequentially contiguous in KF and RT but separated by an intervening residue in pol β . Remarkably, despite these differences, the geometry of the active sites are superposable, suggesting that future structural studies on other polymerases will reveal the same active site geometry and hence the same catalytic mechanism.

At this point, it may be appropriate to consider the issue of evolutionary relationships among the polymerases. Is pol β a distant relative on a single family tree, that is, is it an example of divergent evolution, or does it represent a rare case of convergent evolution? The question is not easily answered. The evident similarity between the palm subdomain in pol β and those of the other polymerases might at first suggest divergence. However the β sheet of the pol β palm has a different topological connectivity, with the strand carrying a key residue, Asp²⁵⁶, inserted into an otherwise antiparallel sheet in the "wrong" direction (Fig. 2). In fact, when one focuses on the geometrically and sequentially invariant portions of

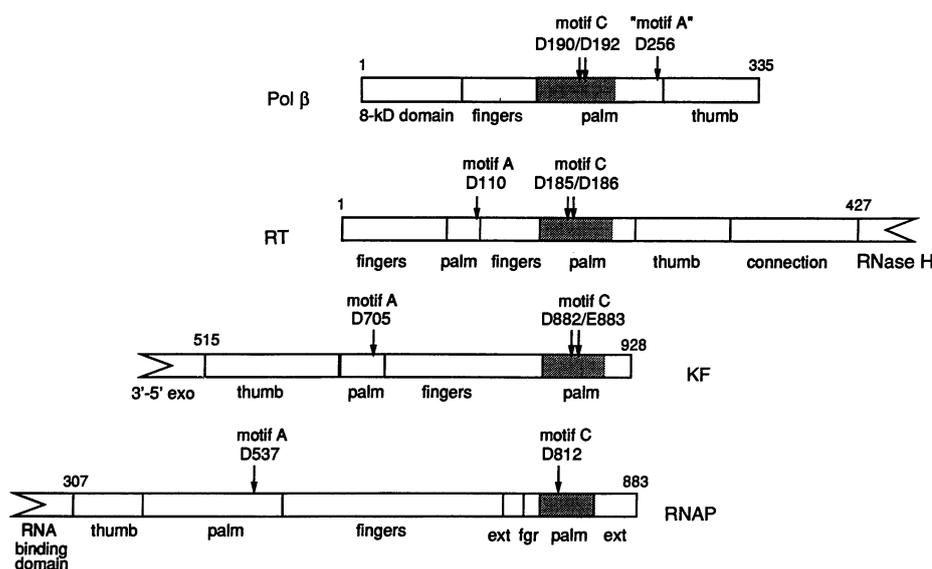


Fig. 3. Locations of the subdomains in the linear sequence of DNA pol β , RT, KF, and T7 RNAP. This figure is a modified version of figure 2 in (14). The sequences are arranged so that motif C of the four polymerases lines up in a vertical column. The shaded portion of the sequence corresponds to the approximately 60-residue segment that is topologically conserved among the four polymerases. In RNAP, fgr stands for fingers and ext stands for extra structure not described by the hand analogy.

the palms of all four polymerases, one is left with a folding motif consisting of just four secondary structural elements, α - β - β - α , shown in gray in Fig. 2. This particular small folding motif is not unusual among many different types of proteins. Moreover, not only do the fingers and thumb subdomains of pol β bear no resemblance to the other polymerase subdomains, but they occur in a different sequential order from the others (Fig. 3). Now it begins to seem that pol β represents a totally unrelated line of descent that has arrived at the same nucleotidyl transfer mechanism independently.

In view of the similarity among the palm folds in KF, RT, RNAP, and pol β , the question arises what, if any, structural features make the palm especially suited for polymerase activity, and whether this fold occurs in other proteins as well. A recent review of α + β structures revealed that polymerase-like palm folds occur in a highly populated group of proteins characterized as two-layered α - β sandwiches (20). These proteins exhibit a wide range of functions, with the palm fold existing as autonomous proteins, or as separate domains of multidomain proteins, or as subdomains within larger proteins.

A computerized search through 278 structures in the Protein Data Bank (29) with the pol β palm (residues 152–262) as a search model yielded 50 matches with rms deviations less than 3.8 Å for between 30 and 50 α carbons. Restricting matches to α - β sandwiches with primarily antiparallel sheets, a solvent accessible β -sheet surface, and two helices on one side of the sheet with matching directionality reduced the number of matches to five (number of α carbons matched and their rms deviation are shown in parenthesis): the activation domain of procarboxypeptidase B (30) (24 α carbons, 3.5 Å); E2 DNA-binding domain (31) (60 α -C, 2.7 Å); glutamine synthetase (32) (60 α -C, 2.4 Å); biotin synthetase (33) (60 α -C, 3.0 Å); and the allosteric domain of aspartate carbamoyltransferase (34) (46 α -C, 3.1 Å). Proteins not included in the search, but later identified by visual inspection include acylphosphatase (35) (42 α -C, 2.6 Å); histidine-containing phosphocARRIER protein (36) (59 α -C, 2.3 Å); L30 ribosomal protein (37) (24 α -C, 2.4 Å); U1 snRNP (38) (54 α -C, 2.5 Å); and the TATA binding protein (39) (50 α -C, 2.5 Å). All the proteins mentioned bind phosphate, nucleotides, or polynucleotides with the exception of the activation domain of procarboxypeptidase B, which may be excluded from the group by its relatively high rms deviation in α -carbon positions.

The α - β sandwiches have been recognized previously as having affinity for phosphate groups. Structural similarities among phosphocARRIER protein, ribonucleoproteins

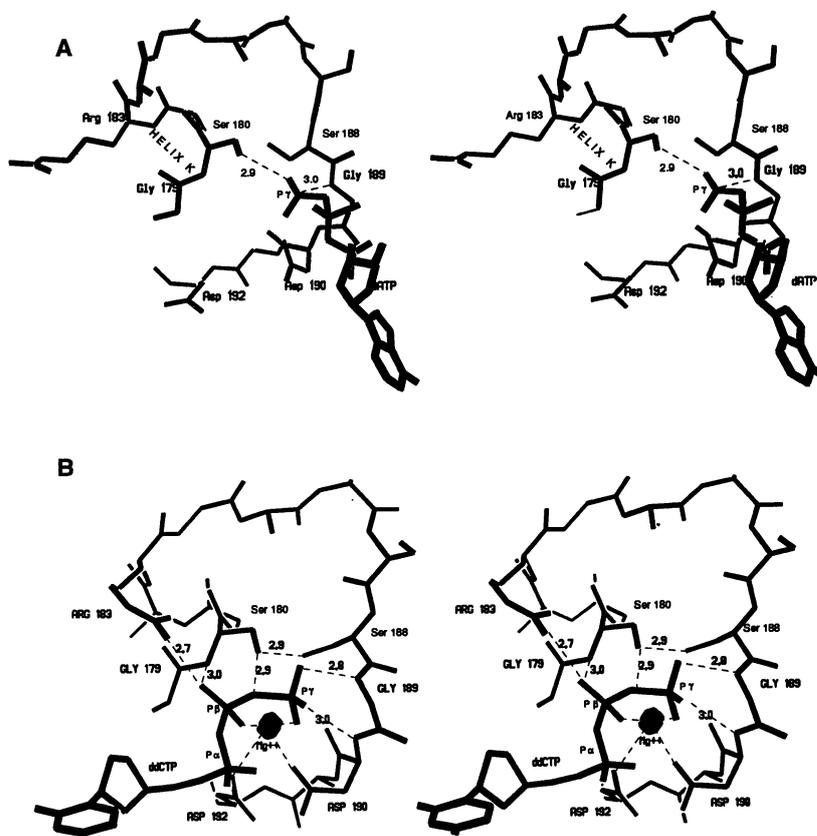


Fig. 4. Comparison of dNTP binding modes in (A) binary dATP complex and (B) ternary ddCTP-*template-primer* complex (5). The binary complex suggests adventitious binding consistent with an incomplete binding site. When the binding site is completed by *template-primer*, binding is productive. *Template-primer* is not shown.

and acylphosphatase suggested that they belong to a new phosphate-binding motif (40). With this latest structural search, it appears that the phosphate-binding motif is actually part of a much larger group of proteins including the polymerases and other nucleotide binding proteins mentioned above. Despite structural similarity, however, the nucleotide binding locus in the α - β sandwich is different in each case. Therefore, we can as yet offer no rationale for the palm's affinity for nucleotides nor for their pattern of occurrence. Nevertheless, this fold is observed in many nucleotide binding enzymes.

Nucleoside triphosphate binding interactions with pol β differ between the ternary *template-primer-ddCTP* complex (5), and the binary dATP complex (Fig. 4). Only the γ -phosphate is similarly bound in both structures. In the binary complex, the nucleoside portion of the dATP projects away from the active site so that the adenine ring may make van der Waals contact with residues from another asymmetric unit. In the ternary complex the ddCTP is held close to the active site by hydrogen bonds with the *template* and van der Waals contacts with the 3'-terminal primer base [specific interactions between pol β and ddCTP are given in (5)]. Presumably these

additional contacts are responsible for fidelity and proximity of P_{α} to the attacking 3'-OH of the primer strand (41). This large difference between dNTP binding in the binary and ternary complexes is in agreement with kinetic studies showing that catalytically productive dNTP binding requires prior *template-primer* binding (42). This difference is also in agreement with the recently reported structure of a binary complex of KF with dNTP, showing nucleotide phosphates bound 6 Å from the conserved aspartic acids Asp⁸⁸² and Asp⁷⁰⁵ (43), where it was also mentioned that such a mode of binding is probably adventitious, the result of an incomplete binding site.

On the basis of the ternary complex structure, we have found that the nucleotide binding fold of pol β contains some of the same structural features found in the classical mononucleotide-binding fold, for example in the kinases (44). Similarity with the consensus sequence of the classical mononucleotide binding fold (45) lies primarily in a glycine-rich loop of pol β which has the same phosphate binding function as the classical fold. In pol β , the γ -phosphate hydrogen bonds to the backbone amide of Gly¹⁸⁹, a residue in the loop (residues 184–189) connecting the NH₂-terminus of β

strand 2 with the COOH-terminus of α helix K (residues 179–183). However, the nucleotide lies on the surface of the antiparallel β sheet, not draped over the end of the parallel sheet as in the classical fold; and although the phosphates lie at the NH₂-terminus of the helix (as in the classical fold), the helix consists of only one turn, making a much weaker dipole interaction with the phosphates. Furthermore, the arrangement of the helix with respect to the β strand is different, with the helical axis lying in the plane of the β sheet. Pol β also lacks the invariant lysine found in the classical fold; however, Arg¹⁸³ similarly hydrogen bonds to the β -phosphate in pol β . The fact that the loop interacts with the β -phosphate in kinases but with the γ -phosphate in pol β supports the concept that the loop interacts only with non-transferrable phosphates. The conserved acidic residues of motif C, Asp¹⁹⁰ and Asp¹⁹² are coordinated by Mg²⁺ (46) which in turn coordinates the β - and γ -phosphates of the nucleoside triphosphate. These aspartates perform a role similar to that of the serine and threonine in the kinases. The role of the third conserved aspartate, Asp²⁵⁶, and the location of a second Mg²⁺ binding site is discussed in (5).

Note added in proof: After this report had been submitted for publication, a description of the 2.3 Å crystal structure of the 31-kD domain of pol β appeared (47). Their 31-kD structure differed from our 31-kD structure in that a cis-peptide bond between Gly²⁷⁴ and Ser²⁷⁵ was not reported. Also, only one Mn²⁺ ion was observed in our 31-kD Mn²⁺ crystal structure, whereas two Mn²⁺ ions, both coordinated to Asp¹⁹⁰, were observed in their 31-kD Mn²⁺ structure.

REFERENCES AND NOTES

- P. Bork *et al.*, *Protein Science* 1, 1677 (1992).
- P. Strauss, *FASEB J. Abstr.* 7, (7), A1291 (1993).
- S. H. Wilson, in *The Eukaryotic Nucleus: Molecular Biochemistry and Macromolecular Assembly*, P. R. Strauss and S. H. Wilson, Eds. (Telford, Caldwell, NJ, 1990), vol. 1, pp. 199–234.
- R. K. Singhal and S. H. Wilson, *J. Biol. Chem.* 268, 15906 (1993).
- H. Pelletier, M. R. Sawaya, A. Kumar, S. H. Wilson, J. Kraut, *Science* 264, 1891 (1994).
- A. Kumar *et al.*, *J. Biol. Chem.* 265, 2124 (1990).
- We had intended to crystallize the entire molecule under these conditions, but apparently an infecting microbial protease cleaved the 39-kD enzyme at a protease hypersensitive hinge connecting the 8-kD and 31-kD domains. This hypothesis is supported by the observation that crystal growth seemed to improve under conditions that favored microbial growth: (i) 25°C rather than 4°C, (ii) in the absence of sodium azide, an antimicrobial agent commonly utilized in protein crystallization experiments, and (iii) when microseeding techniques were used (microseeding probably facilitated contamination by the infecting organisms). The proteolytic step proved necessary for crystallization in space group *P2₁2₁2*, but introduced an uncontrolled complication in reproducing crystallization conditions. When crystals did grow, they usually appeared within 2 weeks after microseed-
- ing. SDS-gel electrophoretic analysis of dissolved crystals showed that they were composed of approximately 90% 31-kD domain and 10% 39-kD enzyme. Because the 8-kD domain was present in low occupancy it was not observed in the electron density map. This unusual packing phenomenon may be explained by the observation that the *P2₁2₁2* crystal lattice interstices appear large enough to accommodate a denatured 8-kD domain in 10% of the asymmetric units scattered throughout the lattice.
- A 2.5 Å data set was collected on a *P2₁2₁2* crystal grown from purified 31-kD domain. This crystal was isomorphous with the crystals grown with proteolyzed 39-kD enzyme. A difference Fourier map revealed only one significant positive peak, near the α -phosphate binding site of the dNTP. This peak probably represents a sulfate anion, made visible by the higher concentration of (NH₄)₂SO₄ present in the crystallization medium.
- J. R. Casas-Finet *et al.*, *J. Biol. Chem.* 266, 19618 (1991).
- R. Prasad, W. A. Beard, S. H. Wilson, *ibid.*, in press.
- A. Matsukage, K. Nishikawa, T. Ooi, Y. Seto, M. Yamaguchi, *ibid.* 262, 8960 (1987).
- T. Raabe *et al.*, *Nature* 353, 229 (1991).
- D. L. Ollis, P. Brick, R. Hamlin, N. G. Xuong, T. A. Steitz, *ibid.* 313, 762 (1985).
- L. A. Kohlstaedt, J. Wang, J. M. Friedman, P. A. Rice, T. A. Steitz, *Science* 256, 1783 (1992).
- A. Jacobo-Molina *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 90, 6320 (1993).
- R. Sousa *et al.*, *Nature* 364, 593 (1993).
- An NMR structure of the 8-kD domain features similar secondary structure assignments. G. Mullen, R. Prasad, S. H. Wilson, *Biochemistry*, in press.
- D. Moras, *Nature* 364, 572 (1993).
- L. S. Beese, V. Derbyshire, T. A. Steitz, *Science* 260, 352 (1993).
- C. A. Orengo and J. M. Thornton, *Structure* 1, 105 (1993).
- There are two ways to superpose the α -carbons of any two palms because of a pseudo twofold rotation axis that passes through the center of the palm and normal to the β -sheet face. Although the rms deviation of α -carbons for the two superposition solutions are approximately equal, only one of the orientations also superposes the conserved aspartic acids of motifs A and C of pol β with the corresponding residues in the other polymerases. This orientation was chosen for all subsequent comparisons. Designation of thumb and fingers in pol β was made so that the thumb lies on the same side of the palm as the thumbs of the other polymerases, given the orientation of the palm defined above.
- M. Delarue, O. Poch, N. Tordo, D. Moras, P. Argos, *Protein Eng.* 3, 461 (1990).
- A. H. Polesky, T. A. Steitz, N. D. F. Grindley, C. M. Joyce, *J. Biol. Chem.* 265, 14579 (1990).
- A. H. Polesky *et al.*, *ibid.* 267, 8417 (1992).
- B. A. Larder *et al.*, *Nature* 327, 716 (1987).
- G. Bonner, D. Patra, E. M. Lafer, R. Sousa, *EMBO J.* 11, 3767 (1992).
- T. Date *et al.*, *Biochemistry* 30, 5286 (1991).
- Only one Mn²⁺ ion was observed in the 31-kD *P2₁2₁2* crystals that had been soaked in the presence of MnCl₂ (Table 1), and its position corresponds to the Mg²⁺ of site B in the ternary template-primer-ddCTP complex as described in (5). It has been shown that Mn²⁺ can substitute functionally for Mg²⁺, the more predominant ion in the cell; Mn²⁺ substitution lowers the Michaelis constant (K_m) for template-primer and increases processivity to 4 to 6 nt inserted per binding cycle. [T. S.-F. Wang and D. Korn, *Biochemistry* 21, 1597 (1982).]
- The 278 protein structures included in the search were selected from the Protein Data Bank so that no pair of proteins have more than 50% sequence homology. These select structures were distributed with the program WHATIF [G. Vriend and C. Sander, *Proteins* 11, 52 (1991)].
- M. Coll, A. Guasch, F. X. Aviles, R. Huber, *EMBO J.* 10, 1 (1991).
- R. S. Hegde *et al.*, *Nature* 359, 505 (1992).
- M. M. Yamashita *et al.*, *J. Biol. Chem.* 264, 17681 (1989).
- K. P. Wilson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89, 9257 (1992).
- J. E. Gouaux, R. C. Stevens, W. N. Lipscomb, *Biochemistry* 29, 7702 (1990).
- A. Pastore, V. Saudek, G. Ramponi, R. J. P. Williams, *J. Mol. Biol.* 224, 427 (1992).
- O. Herzberg *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89, 2499 (1992).
- K. S. Wilson *et al.*, *ibid.* 83, 7251 (1986).
- K. Nagai *et al.*, *Nature* 348, 515 (1990).
- D. B. Nikolov *et al.*, *ibid.* 360, 40 (1992).
- M. B. Swindells *et al.*, *ibid.* 362, 299 (1993).
- The location of dNTP binding was previously thought to involve Lys⁷² in the 8-kD domain, a residue protected from reaction with pyridoxal 5'-phosphate in the presence of template-primer [A. Basu, P. Kedar, S. H. Wilson, M. J. Modak, *Biochemistry* 28, 6305 (1989)]. Later it was found in the electron density map that Lys⁷² forms a subsidiary phosphate binding site along with several other lysines in the 8-kD domain. This site may bind the 5' phosphate of the downstream oligonucleotide in gapped DNA (10).
- K. Tanabe, E. W. Bohn, S. H. Wilson, *Biochemistry* 18, 3401 (1979).
- L. S. Beese *et al.*, *ibid.* 32, 14095 (1993).
- G. Schulz, *Curr. Opin. Struct. Biol.* 2, 61 (1992); M. Saraste, P. R. Sibbald, A. Wittinghofer, *Trends Biochem. Sci.* 15, 430 (1990).
- R. K. Evans, C. M. Beach, M. Coleman, *Biochemistry* 28, 713 (1989).
- No Mg²⁺ ions were visible in the 3.9 Å map of the binary complex with dATP in space group *P2₁*. It is possible that Mg²⁺ ions are present but are not observed in this structure because of their small size and the poor resolution of the map.
- J. F. Davies *et al.*, *Cell* 76, 1123 (1994).
- R. Hamlin, *Methods Enzymol.* 114, 416 (1985).
- A. J. Howard, C. Nielsen, N.-H. Xuong, *ibid.*, p. 452.
- The PCMPDS derivative was prepared by adding 1 μ l of 1 mM PCMPDS to the 10- μ l crystal drop. The crystal was allowed to soak for 3 hours. The K₂PtCl₄ derivative was prepared by adding 1 μ l of 1 mM K₂PtCl₄ to the 10- μ l crystal drop. The crystal was allowed to soak for at least 1 week. Shorter soaks produced no observable difference Patterson peaks. Heavy atom reagents were dissolved in reservoir solution. Methyl mercury bromide, mersalyl acid, and mercurous sulfate produced the same mercury location as PCMPDS, but were less isomorphous.
- B. C. Wang, *Methods. Enzymol.* 115, 90 (1985).
- R. Read, *Acta Crystallogr.* A42, 140 (1986).
- T. A. Jones, *Methods Enzymol.* 115, 157 (1985).
- D. E. Tronrud, L. F. Ten Eyck, B. W. Matthews, *Acta Crystallogr.* A43, 489 (1987).
- P. M. D. Fitzgerald, *J. Appl. Cryst.* 21, 273 (1988).
- A. T. Brünger, *X-PLOR Manual* (Yale Univ. Press, New Haven, CT, version 3.1, 1992).
- P. J. Kraulis, *J. Appl. Cryst.* 24, 946 (1991).
- Supported in part by NIH grants GM10928 and CA17374 (J.K.), NIH grant ES06839 and Welch Foundation Grant H-1265 (S.H.W.), and grants of computer time from the San Diego Supercomputer Center. We thank the personnel of the N. H. Xuong laboratory for aid in data collection, especially C. Nielsen, N. Nguyen, D. Sullivan, and V. Ashford. We thank B. C. Wang for supplying α -carbon coordinates of the palm of bacteriophage T7 RNA polymerase, E. Arnold for supplying α -carbon coordinates for HIV reverse transcriptase, and P. Sigler for supplying coordinates for TATA binding protein. We thank J. Grimsley and W. Woffle for technical assistance in protein preparation. Atomic coordinates for the 31-kD domain, the 31-kD domain complexed with Mn²⁺, the full 39-kD enzyme, and the full 39-kD enzyme complexed with dATP are available from the Protein Data Bank. The coordinates sets are coded 1bpb, 1bpc, 1bpd, and 1bpe, respectively.

28 February 1994; accepted 10 May 1994