Structural basis for the 3′ − 5′ exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism

Lorena S.Beese and Thomas A.Steitz[1]

Departments of Molecular Biophysics and Biochemistry and Chemistry and Howard Hughes Medical Institute, Yale University, New Haven, CT 06511, USA

[1]To whom correspondence should be addressed

The refined crystal structures of the large proteolytic fragment (Klenow fragment) of *Escherichia coli* DNA polymerase I and its complexes with a deoxynucleoside monophosphate product and a single-stranded DNA substrate offer a detailed picture of an editing 3′−5′ exonuclease active site. The structures of these complexes have been refined to *R*-factors of 0.18 and 0.19 at 2.6 and 3.1 Å resolution respectively. The complex with a thymidine tetranucleotide complex shows numerous hydrophobic and hydrogen-bonding interactions between the protein and an extended tetranucleotide that account for the ability of this enzyme to denature four nucleotides at the 3′ end of duplex DNA. The structures of these complexes provide details that support and extend a proposed two metal ion mechanism for the 3′−5′ editing exonuclease reaction that may be general for a large family of phosphoryltransfer enzymes. A nucleophilic attack on the phosphorous atom of the terminal nucleotide is postulated to be carried out by a hydroxide ion that is activated by one divalent metal, while the expected pentacoordinate transition state and the leaving oxyanion are stabilized by a second divalent metal ion that is 3.9 Å from the first. Virtually all aspects of the pretransition state substrate complex are directly seen in the structures, and only very small changes in the positions of phosphate atoms are required to form the transition state. *Key words:* DNA polymerase I/3′−5′ exonuclease/Klenow fragment/three-dimensional structure/X-ray crystallography

Introduction

Biochemical, structural and genetic experiments have demonstrated that the DNA polymerase and 3′−5′ exonuclease activities of *Escherichia coli* DNA polymerase I reside on different domains of its large proteolytic fragment (Klenow fragment) on active sites that are separated by ∼ 33 Å in the crystal structure (Ollis *et al.*, 1985; Freemont *et al.*, 1986; Derbyshire *et al.*, 1988; reviewed in Joyce and Steitz, 1987; Beese and Steitz, 1989). In addition to being spatially separated, the polymerase and exonuclease active sites function independently and have separate binding sites for DNA. Even under conditions where the polymerase active site is occupied by a duplex DNA complex, Klenow fragment can bind a second DNA substrate and carry out exonucleo-lytic cleavage (Catalano and Benkovic, 1989′). In DNA

polymerase III holoenzyme from *E.coli*, the DNA polymerase and 3′−5′ exonuclease activities reside on separate subunits rather than two domains of the same subunit (Maki and Kornberg, 1985; Scheuermann and Echols, 1984). Since these two activities can be considered separately, we shall focus here on the structural details of the 3′−5′ exonuclease active site and its enzymatic mechanism.

That the substrate for the exonuclease reaction is at least partially single-stranded rather than duplex DNA was suggested by Brutlag and Kornberg (1972) from the temperature dependence of the exonuclease reaction. Further, the crystal structure of the Klenow fragment co-crystallized with an 8 bp duplex DNA showed a single-stranded oligonucleotide bound to the exonuclease site (Steitz *et al.*, 1987; Freemont *et al.*, 1988) and thus established directly that the exonuclease active site melts duplex DNA and binds single- rather than double-stranded DNA. Subsequently, results of exonuclease digestion of crosslinked duplex DNAs suggested that the Klenow fragment requires between 4 and 5 bp of the primer strand to melt out from a duplex DNA for the exonucleolytic removal of nucleotides from the primer terminus (Cowart *et al.*, 1989). Since the exonuclease active site is able to denature DNA at its 3′ terminus and bind single-stranded DNA, it must provide more favorable stabilizing interactions with the single-stranded DNA than those that stabilize duplex DNA.

These observations suggest that one way in which *Pol*I exerts control over the excision of a mismatched base pair could be the decreased stability of such a duplex, making it a poorer substrate for additional DNA synthesis and increasing the probability of forming a single strand in the exonuclease site (Steitz *et al.*, 1987). The binding of single-stranded DNA to the exonuclease active site would result in excision of the terminal nucleotide whether or not it had been correctly incorporated, consistent with the finding that 2−3% of correctly incorporated nucleotides are excised (Fersht *et al.*, 1982; Joyce, 1989).

Significant insights into the specificity and reaction mechanism of the 3′−5′ exonuclease reaction have been obtained from earlier structural studies of both product and substrate complexes (Ollis *et al.*, 1985; Freemont *et al.*, 1988). The deoxynucleoside monophosphate binds to the smaller N-terminal domain of the Klenow fragment in the same positions as the 3′ terminal nucleotide of a single-stranded substrate. In the absence of substrate or product, the exonuclease domain has one firmly bound divalent metal ion (site A) while the deoxynucleoside monophosphate complex shows a second divalent metal ion (site B) bound ∼4 Å from site A. Divalent metal ions (Mg^{2+}, Mn^{2+} or Zn^{2+} but not Ca^{2+}) are essential for exonuclease activity (Lehman and Richardson, 1964; Kornberg, 1980) and their removal by EDTA or by mutations results in the loss of dNMP binding to the crystal and the loss of exonuclease activity in solution (Derbyshire *et al.*, 1988). Of particular importance for the DNA substrate complex described here

is the mutant protein D424A, which shows no measurable exonuclease activity but a nearly identical structure for the dNMP complex, except that metal ion B no longer binds. Not only does this imply an important catalytic role for metal ion B, but it also allows the preparation of a stable complex with a single-stranded DNA substrate.

The refined structures of the substrate and product complexes with the 3'−5' exonuclease active site described here provide details that extend and support a proposed two metal ion mechanism for this hydrolytic phosphoryl transfer reaction (Freemont *et al.*, 1988). We conclude that the chemical catalysis of the hydrolytic phosphoryl transfer reaction is promoted by the two metal ions, while the roles of the protein residues are to bind and position the two metal ions, the single-stranded DNA and the attacking water molecule in the relative orientation that is required to make this catalytic mechanism efficient.

Results

The refined structures of substrate and product complexes reveal an elaborate network of H-bonds and water molecules
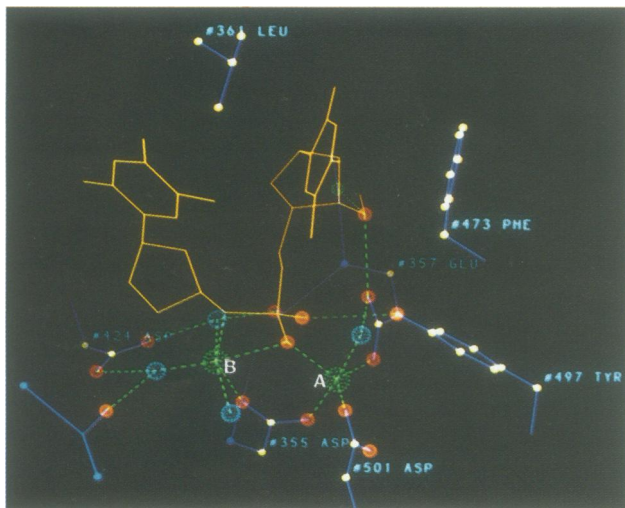

Fig. 1. Overview of the 3'−5' exonuclease active site showing a bound deoxydinucleotide. The positions of the sugar and base of the penultimate nucleotide are determined from the tetranucleotide complex, while the 3' terminal nucleotide and the remaining coordinates are from the dTMP complex. The side chains of residues interacting with the metal ions and dTMP are shown. The two binding sites for divalent metal ions are labeled A and B and are 3.9 Å apart. Metal ligand distances are between 1.9 and 2.2 Å and H-bond distances are between 2.7 and 3.0 Å.
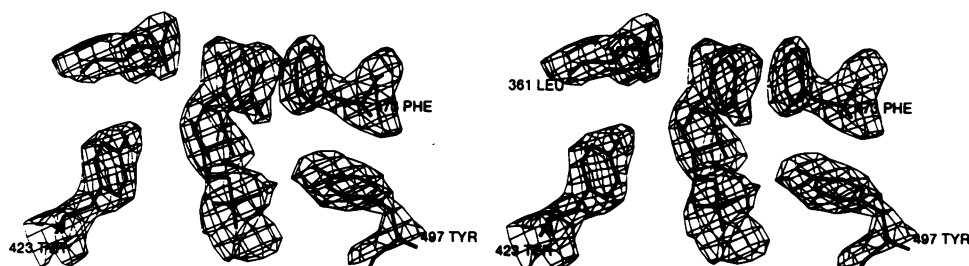
in the active site around the bond to be cleaved. These structures are similar in overall appearance to the exonuclease domain reported in Ollis *et al.* (1985), but differ substantially in detail, particularly in the active site. We describe here two structures: the first, refined at 2.6 Å resolution, is the wild-type protein complexed with product, deoxynucleoside monophosphate; and the second, refined at 3.1 Å resolution, is a mutant protein (D424A) that is deficient in exonuclease activity complexed with a single-stranded DNA substrate.

Product complex

The product dTMP is bound to a single site with its phosphate and ribose moieties nearly buried in a hydrophobic binding pocket, and its phosphate interacting with two divalent metal ions that are separated by 3.9 Å. The conformation of the nucleotide is approximately that found in B-DNA, having a C3' endo sugar pucker with the thymidine base in an anti, periplanar conformation. Modification of any of the 5' phosphate oxygen atoms abolishes dNMP binding, implying that all available phosphate oxygen atoms make important interactions (Kornberg, 1980). These phosphate oxygens interact with the two metal ions, and the hydroxyl of Tyr497 (Figure 1). The 3' OH group is buried and H-bonded to the backbone amide proton of Thr358 and the carboxylate of Glu357, which may explain the failure of dideoxynucleoside monophosphates to bind (Atkinson *et al.*, 1969). There are no specific H-bonding interactions between the thymidine base and the protein, as would be expected because of the ability of the dNMP binding site to bind any of the four bases or many modified bases. Instead, the base is held by surrounding hydrophobic interactions, on one side by the side chain of Leu361 and the other by the side chain of Phe473 which stacks against the 3' terminal base (Figure 2). The edge of the base is accessible to solvent and a water molecule is observed to bind to the O2 of thymidine.

Divalent metals

In the complex between wild-type protein and nucleoside monophosphate there are two bound metal ions separated by 3.9 Å. One metal ion (site A) is coordinated to the protein by the carboxylate groups of Asp355, Glu357 and Asp501, while the 5' phosphate of dNMP provides a fourth ligand (Figure 3). Mn^{2+}, Mg^{2+} or Zn^{2+} ions can bind to this site at full occupancy in the crystal even in the absence of dNMP. A difference electron density map calculated at 2.6 Å resolution using the observed and calculated amplitudes as coefficients (Figure 4) contains density consistent with a water molecule (or hydroxide ion) binding as a fifth ligand

Fig. 2. Electron density superimposed on refined coordinates illustrating Phe473 stacking against the base of dTMP. The electron density map was calculated using Fourier coefficients ($2F_o - F_c$). A 1.6σ contour level is shown.

in a distorted tetrahedral geometry. This water molecule is coordinated by the Zn^{2+} ion, the α-carboxylate of Glu357 and the hydroxyl of Tyr497. The pentacoordinated metal ion, showing a distorted tetrahedral geometry, is similar to that observed in a number of zinc enzymes.

A second divalent metal ion (site B), located between the dNMP phosphate and the carboxylate of Asp424, is observed only when dNMP is bound. It is octahedrally coordinated by the carboxylate of Asp355, two of the 5′ phosphate oxygens of dNMP, and three water molecules, each of which is in turn bound to protein. Two of these water molecules form H-bonds to the carboxyl oxygens of Asp424 and to backbone amides (Figures 1 and 5). This elaborate network of H-bonds suggests why the mutation of Asp424 to Ala disrupts metal B binding. The positions of water molecules in the exonuclease active site were checked at the end of the refinement by calculating a series of difference maps in which the water molecules were omitted from the calculated structure factors (Figure 4).

Although we do not know the identity of the metal ions that are bound *in vivo*, in the presence of both Zn^{2+} and Mg^{2+} metal ions, Zn^{2+} binds to site A and Mg^{2+} to site B. This is established by the presence of a large negative peak at site B in a difference electron density map calculated between diffraction data collected from a crystal soaked in a solution containing both 20 mM $MgSO_4$ and 1 mM $ZnSO_4$ and a crystal containing 1 mM $ZnSO_4$ alone. This shows that in the crystal soaked in both metal ions, the site A metal ion is Zn^{2+} and the site B metal ion is Mg^{2+}. The observation that the Zn^{2+} ion binds preferentially to site A

and the Mg^{2+} ion to site B is consistent with the preferred coordination geometry of the two metal ions, tetrahedral for zinc and octahedral for magnesium ions, that is exhibited by sites A and B respectively.

Single-stranded DNA complex

The three nucleotides at the 3′ end of the tetranucleotide fit into a buried hydrophobic pocket that contains the two divalent metal ions at the center of the exonuclease active site (Figure 6). The 3′ terminal nucleotide binds in a nearly identical fashion to deoxynucleoside monophosphate, except that its 5′ phosphate lies in a slightly altered orientation (Figure 3). The 5′ terminal nucleotide binds in a shallow groove on the surface that extends out of the pocket to the interface with the polymerase domain (Figure 6). The binding site spans ~18 Å. This binding site appears to be 'designed' to bind single-stranded DNA rather than duplex since it is difficult to model build a complementary DNA strand in this groove without creating numerous close contacts with the protein.

The binding of tetranucleotide produces only a few significant changes in the protein structure, largely confined to side chains in contact with the nucleic acid. The r.m.s. deviation in protein coordinates between the DNA and the dTMP complexes is 0.35 Å, which is about the expected error in the coordinates.

Protein residues that form both the pocket and groove are accessible to solvent in the apo enzyme and bind several water molecules that are displaced upon the binding of nucleotides. The displacement of these water molecules contributed to difficulty in interpreting the position of the third and fourth nucleotides in the difference electron density map between the tetranucleotide complex and the apo enzyme (Freemont *et al.*, 1988). This problem is circumvented by subtracting the protein model (without these water molecules) from the complex. The fit of the refined
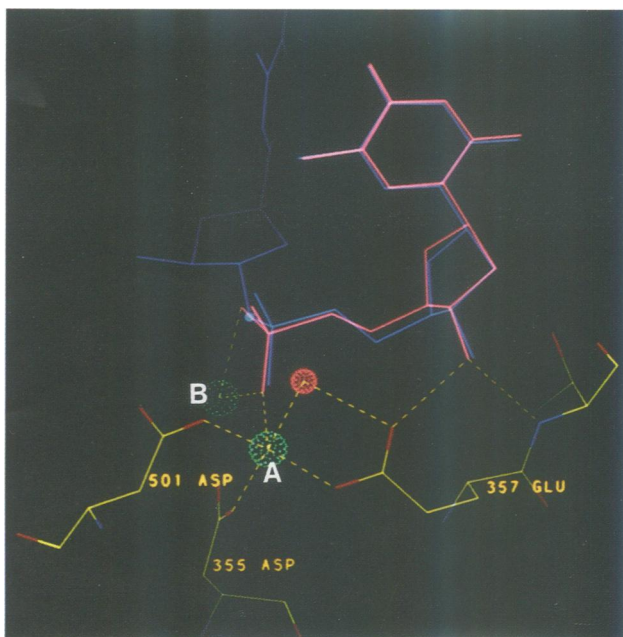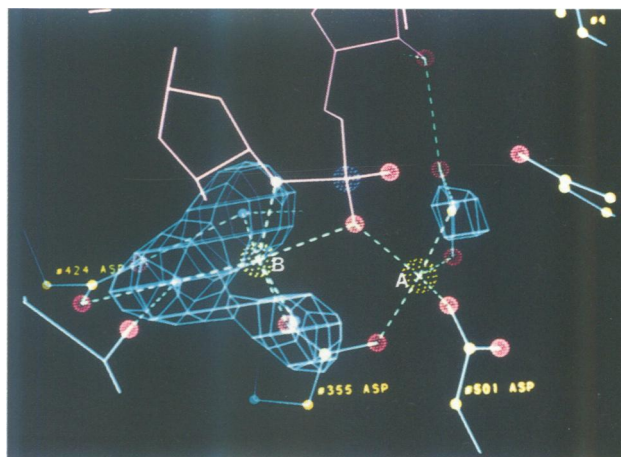

Fig. 3. Superposition of the 3′ terminal dinucleotide from the tetranucleotide (blue) on nucleoside monophosphate (red). The position of the terminal nucleotide is almost identical in the two cases. Metal ions are represented by green spheres. Metal ion A (Zn^{2+}) has a distorted tetrahedral coordination geometry with a water molecule forming a fifth ligand (red sphere). The distances between Zn^{2+} (A) and its ligands are 2.0, 2.0, 2.0, 2.1 and 2.1 Å to D355, D501, E357, water (hydroxide) and phosphate oxygen respectively. The distances to metal ion B are 2.2, 2.1, 2.4 Å to D355, phosphate oxygen and 3′ oxygen leaving group respectively. The ribose of the terminal nucleotide is oriented and bound by two hydrogen bonds to 3′ sugar hydroxyl, one to backbone amide and the other to Glu357.

Fig. 4. Difference electron density map ($F_o - F_c$) at 2.6 Å resolution showing the water molecules that interact with the metal ions calculated with the water molecules omitted from the calculation. The map is shown at 3.5σ contour. To minimize the bias of this map, 50 cycles of positional refinement with the tetranucleotide omitted were done prior to calculation of the map and SigmaA weight (Read, 1986), and were applied to the structure factors used in the map calculation. Note particularly the density for the fifth ligand to metal ion A, which we believe is a water molecule or hydroxide ion that is opposite to the phosphodiester bond being cleaved and oriented to attack the phosphorus.
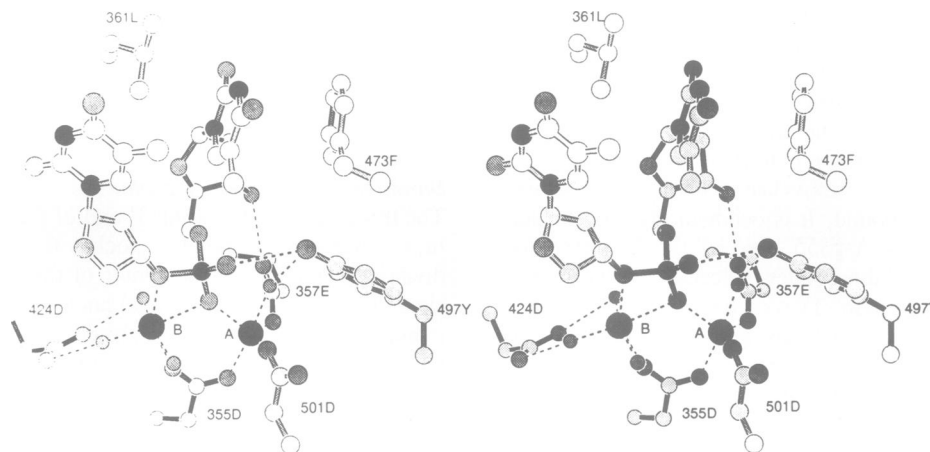
Fig. 5. A stereoscopic representation of a dinucleotide bound at the 3'−5' exonuclease active site. The coordinates of the dinucleotide were taken from those of the tetranucleotide complex while the coordinates for the protein were taken from those of the dTMP complex. The two metal ions, A and B, are represented by large black balls while the phosphorus is represented by a smaller black ball. The atoms of carbon, oxygen and nitrogen are represented by increasingly darker shades of grey. Water molecules are shown as smaller grey spheres. Some of the H-bonding interactions seen in this complex are shown as dashed lines. Note that the coordination of metal ion A is a distorted tetrahedral coordination with a water molecule forming the fifth ligand, while that around metal ion B is a distorted octohedral coordination.
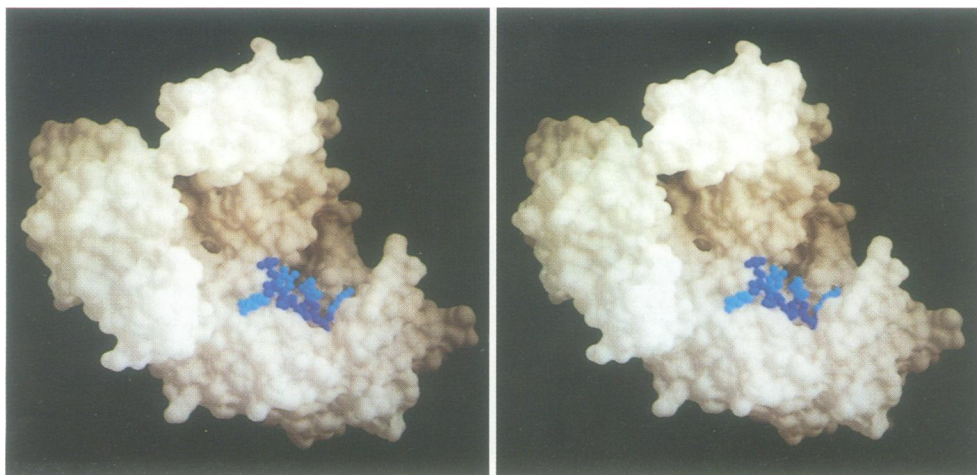

Fig. 6. A stereo drawing of a solvent accessible surface representation (Connolly) of the Klenow fragment with a ball-and-stick representation of single-stranded thymidine nucleotides (dark blue, sugar-phosphate backbone; light blue, bases) bound in a cleft leading from the exonuclease active site. The 3' terminal nucleotide cleaved during the reaction (rightmost in view shown) is bound in a buried pocket. To the left and above the single-stranded DNA is the cleft that contains the polymerase active site.

nucleotide to a difference electron density calculated using $[F_{obs}(complex)_{hkl} - F_{calc}(protein)_{hkl}]$ as coefficients is shown in Figure 7.

The four nucleotides are bound to the enzyme by hydrophobic interactions and hydrogen bonds that are independent of the identity of the base. The 3' terminal two nucleotides are the most well ordered and have refined to temperature factors of 20 and 25 Å^2. Hydrophobic interactions of three protein side chains with the two terminal bases hold these nucleotides tightly. The side chain of Leu361 is wedged between the 3' terminal and penultimate bases (Figure 6), thereby disrupting the base stacking of the nucleotides. Leu361 is part of a main chain that had been rebuilt during refinement and extends further between the bases than previously observed (Freemont *et al.*, 1988). Tyr431 makes van der Waals contact with the sugar residues of the terminal two nucleotides. The 4' oxygen of the 3' penultimate nucleotide hydrogen bonds to the δ amino of

Asn420. As in the dTMP complex, Phe473 stacks against the 3' terminal base.

The third nucleotide from the 3' terminus interacts with His660, which has two conformations (Figure 8). In one conformation it stacks against the base, and in the other it could form a hydrogen bond with a 5' phosphate and to Thr672. The latter conformation is observed in the absence of bound nucleotides. Met443 makes van der Waals contact with the sugar of the third nucleotide, whose base hydrogen bonds to Ser658 through the keto oxygen at the 4 position.

The sugar-phosphate backbone of the single-stranded DNA interacts with three amino acid side chains, metal A and presumably also with metal B. In order to prevent substrate hydrolysis, the complex is made with the mutant protein D424A, which does not bind a metal ion in site B. If we assume that metal ion B binds to the substrate complex with the wild-type protein as it does in the product complex, then the phosphate oxygens of the 3' terminal nucleotide would
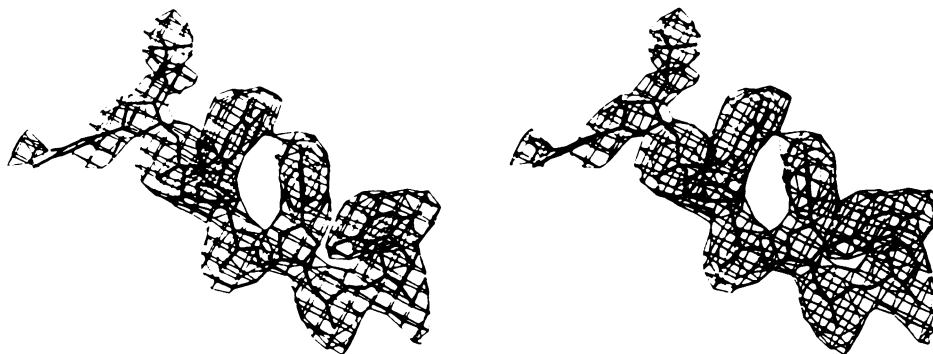
Fig. 7. Refined positions of dT$_4$ superimposed on difference electron density map (F_o − F_c) with the coordinates for the nucleotide omitted from the calculation of the phases and calculated structure factors (F_c). A 2.5σ contour level of the electron density is shown. To minimize the bias of this map, 50 cycles of positional refinement with the tetranucleotide omitted were done prior to calculation of the map and SigmaA weight (Read, 1986), and were applied to the structure factors used in the map calculation. Nucleotides are in an unstacked, extended conformation with sugar and base orientations consistent with those observed in B-form DNA.
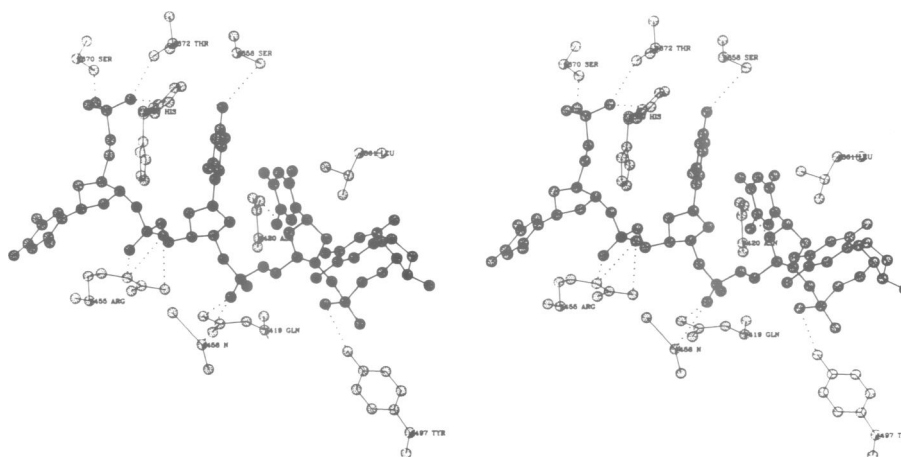

Fig. 8. Protein–DNA interactions. The 3′-terminal two nucleotides of single-stranded DNA are stabilized by hydrophobic interactions between Phe473 (not shown), Leu361 and Tyr431 (not shown) and hydrogen bond between Glu420 and O4′ of sugar. His660 has two conformations: one stacking against the third base, and the other forming a H-bond to the 5′ phosphate and to Thr672 (not shown). The latter conformation is observed in the absence of bound nucleotides. Residues Gln419, Arg455, Tyr497, Ser670, Thr672, His660 and the main chain amide of Met458 form H-bonds with the sugar-phosphate backbone of DNA.

interact with the two metals and Tyr497. Non-esterified phosphate oxygens from the penultimate nucleotide interact with a main chain amide of Met458 and also Gln419. The third phosphate group from the 3′ terminus forms an ion pair interaction with the positively charged guanidinium group of Arg455.

The 5′ terminal nucleotide is accessible to solvent and is quite disordered. Its only interactions with the protein are through H-bonds between its two non-esterified phosphate oxygens and Ser670 and Thr672. The base is not stabilized by the protein, and as a consequence its position is not well determined. From the electron density alone one cannot rule out the possibility that the 5′ terminal nucleotide binds with its base pointing toward Ser670 and its phosphate is disordered. However, the conformations shown in Figures 6−8 agree with what is observed in the structure co-crystallized with the duplex DNA octamer (L.S.Beese and T.A.Steitz, unpublished). The buried single-stranded binding site ends at the sugar of the third nucleotide and is consistent with the observation that oligonucleotides appear to be less well ordered after three nucleotides (Freemont *et al.*, 1988).

The conformation of the single-stranded DNA is extended

and the bases are splayed rather than stacked, particularly those of the 3′ terminal nucleotides. Sugars have C3′ endo sugar pucker and the bases are in an anti-conformation. There are not many known structures of single-stranded DNA for comparison. Although the bases determined from the crystal structure of sodium thymidylyl-5′,3′-thymidylate-5′-hydrate (pTpT) (Camerman *et al.*, 1976) are splayed rather than stacked, many of the torsion angles are quite different, and thus the structures cannot be superimposed. Another example of splayed nucleotides is observed in the structure of the glutaminyl tRNA synthetase−tRNA complex (Rould *et al.*, 1989) where Leu136 is responsible for disrupting base stacking between A72 and G2 in the acceptor stem of the tRNA. Again, however, the position of leucine with respect to the nucleotides in the two complexes cannot be superimposed.

Discussion

Single-stranded DNA binding and editing

The function of the 3′−5′ exonuclease activity in DNA polymerase I is to remove incorrectly incorporated nucleotides. As expected, we see no interactions between
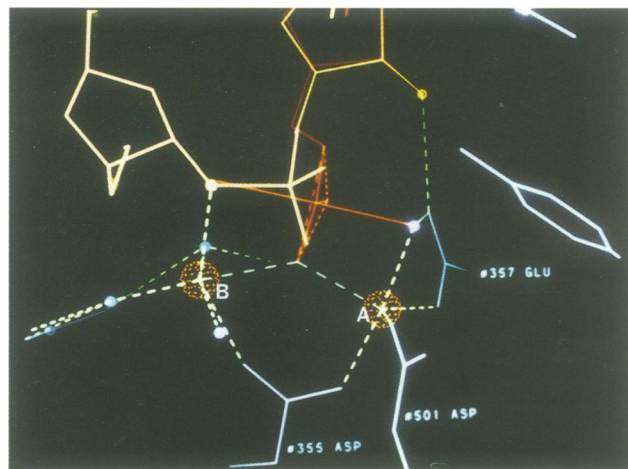
Fig. 9. A model-built pentacoordinate intermediate (in orange) superimposed on the terminal two nucleotides of the tetramer complex. The transient pentacoordinate species was constructed by holding the position of the oxyanion of the leaving group fixed and positioning the phosphorus and its associated equatorial oxygens midway between the two metals. The attacking water molecule was moved 0.5 Å from that observed in the complex with the nucleoside monophosphate product to be equivalent to the P—O distance of the leaving group. The P—O bond lengths are 1.9 Å for apical oxygens and 1.5 Å for the others in a plane with the phosphorus.

Fig. 10. The proposed transition state of the two metal ion enzymatic mechanism for the 3'—5' exonuclease reaction. Metal ion A on the right is proposed to facilitate the formation of an attacking hydroxide ion whose lone pair electrons are oriented towards the phosphorus by interactions with the metal ion, Tyr497 and Glu357. Metal ion B is hypothesized to facilitate the leaving of the 3' hydroxyl group and stabilization of the 90° O—P—O bond angle between the apical and equatorial oxygen atoms.

the protein and the bases that are sequence specific, but rather we see an extended binding site that makes numerous interactions with a single-stranded tetranucleotide. How, then, does this editing active site that lies some 30 Å from the polymerase site of misincorporation (Ollis *et al.*, 1985) exert selective bias in removing mismatched nucleotides with higher frequency than correctly incorporated nucleotides?

A competition model has been proposed in which any factor that destabilizes the DNA duplex (such as mismatched base-pairs, base composition or temperature) would increase the ratio of exonuclease to polymerase activity and thus would lead to a higher rate of excision of mismatched base-pairs compared with correctly paired bases (Steitz *et al.*, 1987; Joyce and Steitz, 1987; Freemont *et al.*, 1988). Additionally, formation of a mismatched base-pair at the polymerase active site could result in a decreased rate of its translocation and subsequent polymerization and/or in an increased rate of its dissociation from the polymerase active site, both of which would increase the exonuclease rate relative to that of the polymerase (Freemont *et al.*, 1988).

Particularly prominent among those interactions between the protein and the single-stranded DNA that would compensate for the loss of base stacking interactions in duplex DNA that occurs upon formation of single-stranded DNA are those provided by the side chains of Leu361 and Phe473. In particular, Phe473 provides a stacking interaction with the 3' terminal base that is not present in duplex DNA, suggesting that it might play a particularly significant role in denaturing the duplex DNA as well as orienting the 3' terminal nucleotide. While the L361A mutant protein showed a 25-fold reduction in activity on duplex DNA when compared to wild type, it showed only a 2- to 3-fold decrease on single-stranded DNA (Derbyshire *et al.*, 1991) implying an important role for the Leu side chain in stabilizing single-stranded versus duplex DNA. Unfortunately, the F473A mutant protein does not have enough activity to assay on single-stranded DNA.
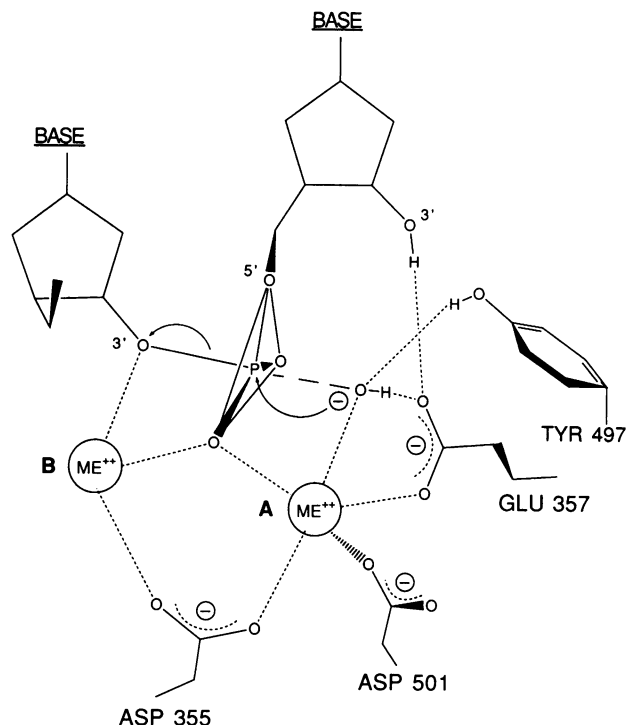
A two metal ion phosphoryltransfer mechanism

The refined structures of the Klenow fragment complexed with a substrate and product at the 3'—5' exonuclease active site not only place constraints on the possible catalytic mechanisms, they now also very strongly support and extend the mechanism proposed earlier on the basis of lower resolution, unrefined structures (Freemont *et al.*, 1988). An intimate and accurate view of a pretransition state substrate complex is possible for this enzyme, in contrast with previous studies of hydrolytic enzymes, because the exonuclease-deficient mutant protein, D424A, produces no perturbation of the protein structure and forms stable substrate complexes (Derbyshire *et al.*, 1988). This allows the direct observation of substrate atoms on both sides of the bond to be cleaved.

The mechanism of the 3'—5' exonuclease reaction proposed by Freemont *et al.* (1988) is shown in Figure 10 with the greater detail and some modifications allowed by the refined structures. Several aspects of this mechanism that previously were only surmised are now seen directly in the refined pretransition state complex. One non-esterified oxygen of the scissile phosphate lies within bonding distance of both metal ions and between them, rather than interacting only with metal ion A as previously thought. Furthermore, in a 2.6 Å resolution electron density map of the dNMP complex, some electron density corresponding to a water molecule (or hydroxide ion) is seen bonded to metal ion A. It is directly opposite the phosphorus from the 3' oxygen of the leaving sugar, exactly appropriate for the attacking hydroxyl ion; Gupta and Benkovic (1984) have shown that hydrolysis results in inversion of configuration at the 3'

terminal phosphorus, implying an associative in-line mechanism with the hydroxyl attacking from a position opposite the 3' hydroxyl leaving group. The putative attacking hydroxyl is seen to be interacting with metal ion A, which may be functioning as a Lewis acid by displacing a proton from water. It is also H-bonded to the metal ion A ligand Glu357, which must be a H-bond acceptor, and the OH of Tyr497, which could only be a H-bond donor to a hydroxide (or an acceptor to water). Thus, the metal ion, the hydroxyl group of Tyr497 and the carboxylate group of Glu357 serve to position exactly the remaining lone pair electrons of the hydroxyl ion towards the phosphorus and the O−P bond it is to attack (Figures 3 and 10).

Only small changes in the positions of the scissile phosphate atoms of the substrate from their observed pretransition state positions are required to form the pentacovalent transition state that is expected in phosphoryl transfer reactions of this type (Figure 9). A hypothetical transient pentacoordinated species was constructed by holding the refined position of the oxyanion of the leaving group fixed and positioning the plane of the phosphorus and three equatorial oxygens midway between the two metal ions. The P−O bond lengths obtained in so doing are 1.9 Å for apical oxygen atoms and 1.5 Å for the others.

What generates the attacking hydroxide ion? There are only three moieties interacting with the hydroxide. Since Glu357 is also a metal ligand, it presumably would not function as a general base but rather serves to orient the hydroxide ion. Although in principle an unprotonated Tyr497 might function as a general base, this possibility is ruled out by the results of mutagenic studies; the mutant protein Y497F shows a very similar pH rate profile to wild type and a modest reduction in activity (Derbyshire et al., 1991). The apparent reduction in activity of the wild-type enzyme but not of Y497F at pH values > 10 is consistent with the phenolic hydroxyl of Tyr497 binding the hydroxyl ion as a proton donor as shown in Figure 10.

This leaves the possibility that the divalent metal ion in site A promotes the formation of the hydroxide ion. The mechanisms of several zinc metalloenzymes have been proposed to be facilitated by the formation of a zinc hydroxide: carboxypeptidase A (Christianson and Lipscomb, 1988, 1989; Kim and Lipscomb, 1990), carbonic anhydrase (Liljas et al., 1972; Lindskog and Coleman, 1973), thermolysin (Holmes and Matthews, 1982; Matthews, 1988) and alkaline phosphatase (Coleman, 1987; Wyckoff, 1987). In each case, crystal structures of substrate analog and inhibitor complexes show three protein and one substrate ligand forming distorted tetrahedral geometry around the metal ion, and a putative hydroxide ion is seen in some cases to be forming a fifth ligand to the metal. We suggest that the metal ion in site A is promoting the formation of the attacking hydroxide ion, while Tyr497 and Glu357 are serving to orient it in the appropriate position for attack.

What is the role of the second metal ion in site B? Most likely this metal ion may function to stabilize the transient pentacovalent species and/or to facilitate the leaving of the 3' oxyanion from an apical position. Herschlag and Jencks (1987) concluded from a series of nucleophilic displacement reactions at phosphorus that divalent metal ions can serve to stabilize an oxyanion leaving group. For phosphomono-esters, divalent metal ions render phospho-group transfer less sensitive to the nucleophilicity of the attacking group while

increasing the overall rate. In addition, the metal ion at site B could be stabilizing the transient pentacovalent species in a manner analogous to that observed with certain organophosphates. For example, the rate of hydrolysis of methyl ethylene phosphate is some 10^6 times faster than that of the trimethyl analog (Westheimer, 1968), presumably because the ethylene bridge strains the O−P−O bond angle to 99°, 10° less than the tetrahedral angle (Steitz and Lipscomb, 1965). Both the structure of the pretransition state complex (Figures 1 and 5) and the model-built pentacovalent intermediate (Figure 9) are consistent with the possibility that metal ion B performs a function analogous to that played by the bridging ethylene group in methlethylene phosphate; it may reduce the energy required to form the 90° O−P−P bond angle between apical and equatorial oxygens in the transition state.

We conclude, then, that the roles of protein residues in this two metal ion mechanism are to bind and orient the two metal ions, the single-stranded DNA and the attacking water molecule in a manner that is essential for catalysis; the catalysis of the hydrolytic phosphoryltransfer reaction is promoted by the two metal ions that serve to facilitate hydroxide ion formation and stabilize the transition state. This mechanism derived from chemical and structural considerations appears to be entirely consistent with the results of site-directed mutagenesis (Derbyshire et al., 1988; Derbyshire et al., 1991). Asp424 forms part of the binding site for metal ion B through two water-mediated H-bonds. Replacement of Asp424 with Ala results in the loss of metal ion B binding and activity, while Glu at this position retains considerable activity, presumably by interacting directly with metal ion B in a less optimal way. The only other mutations that reduce exonuclease activity below the detection limit are mutations of two other metal ion ligands, Asp355 and Asp501. Changes in amino acid side chains that are in contact with the substrates have significant, but more modest effects on the exonuclease activity (Derbyshire et al., 1991). The carboxylate of Glu357 functions to bind and orient metal ion A, the 3' OH of the terminal nucleotide and the attacking hydroxide ion. Tyr497 also orients the attacking water molecule. Phe473 orients the base of the terminal nucleotide, and along with Leu361 provides hydrophobic stabilization of the single-stranded substrate.

The chemical basis of the proposed two metal ion mechanism appears very similar to that of the mechanism for each step of the two-step reaction mechanism proposed (Wyckoff, 1987) for E.coli alkaline phosphatase. This enzyme also contains two bound divalent metal ions (zinc) that are 3.9 Å apart and interact with a phosphate monoester involved in the phosphoryl transfer reaction catalyzed by this enzyme (Kim and Wyckoff, 1991). The orientation of the phosphate relative to the metal ions is strikingly similar in both enzymes (L.S.Beese, E.Kim, H.W.Wyckoff and T.A.Steitz, unpublished observation). Since neither the overall structure nor the residues at the active site of alkaline phosphatase bear any resemblance to the small domain Klenow fragment, the apparently common chemical mechanism presumably results from convergent evolution.

Indeed, this two metal ion mechanism for catalysis of phosphotransfer reactions may be very primitive, since it may only require the proper orientation of the metal ions relative to the substrates. It has been proposed that RNA enzymes involved in reactions leaving a 5' phosphate and

in splicing may use this two metal ion mechanism of phosphoryl transfer, which would not require any RNA functional groups to be directly involved in catalysis of bond formation and breaking (Freemont *et al.*, 1988). Further, since it is widely believed that the polymerase reaction itself may have been catalyzed by RNA enzymes before the evolution of proteins, it will be interesting to see whether that phosphoryl transfer reaction is also catalyzed by a two metal ion mechanism.

The structure and mechanism of the Klenow fragment exonuclease may be applicable to the 3′−5′ exonucleases of other DNA polymerases whose three-dimensional structures are not yet known. Residues at the 3′−5′ exonuclease active site appear to be conserved among many prokaryotic and eukaryotic polymerases (Joyce *et al.*, 1986; Spicer *et al.*, 1988; Bernad *et al.*, 1989; Leavitt and Ito, 1989), although sequence similarities that might suggest homologies between the Klenow fragment and the eukaryotic polymerases are at best very weak. Recent site-directed mutagenesis studies along the lines of Derbyshire *et al.* (1988) have been undertaken on ϕ29 DNA polymerase (Bernad *et al.*, 1989), and have been interpreted to show the importance of certain carboxylate acid residues for binding metals.

Materials and methods

Crystallographic data

Crystals were grown in space group $P4_3$, cell dimensions 103.4 × 103.4 × 86.2 Å, with one molecule per asymmetric unit as described in Ollis *et al.* (1985). Crystals were transferred from growth buffer (1.4 M citrate buffer, pH 5.8) to 70% saturated ammonium sulfate, 66 mM PIPES buffer, pH 7.0, with 2 mM $ZnSO_4$ and 14 mM dTMP. Data were collected using a Mark II two-dimensional position-sensitive area detector (Xuong *et al.*, 1985) to 2.5 Å resolution at −20°C. The merging R-factor = $(\Sigma|I_{i,n} - <I_n>|)/\Sigma<I_n>$ (where $I_{i,n}$ is the ith observation of reflection n) for the data set was 0.051. There is a sharp falloff in data quality past 3.3 Å resolution. All reflections between 8 and 3.3 Å were included in the refinement, but at resolutions higher than 3.3 Å only those reflections where $F_{obs} > 2.5\sigma$ were used. The data were 97% complete between 20 and 3.3 Å, but only 63% complete between 3.3 and 2.6 Å. The resulting data set used for refinement was 78% complete between 20 and 2.6 Å. Crystals of the exonuclease minus mutant (Asp → Ala424) of Klenow fragment (Derbyshire *et al.*, 1988) were grown similarly and were soaked in buffer containing 70% saturated ammonium sulfate, 66 mM PIPES buffer, pH 7.0, with 1 mm deoxythymidine tetranucleotide p(dT$_4$) and 20 mM MgSO$_4$. Data were measured to 2.9 Å with a merging R-factor of 0.055. Data were 90% complete to 3.1 Å resolution using all data with $F_{obs} > 2.5\sigma$.

Refinement of the dTMP complex

Refinement was begun using the least-squares refinement program PROLSQ (Konnert and Hendrickson, 1980; Hendrickson, 1985) with frequent rebuilding into the electron density using the interactive graphics program FRODO (Jones, 1985). The entire molecule was rebuilt using difference electron density $(2F_o - F_c)$ and $(F_o - F_c)$ maps with SigmaA modified coefficients (Read, 1986) omitting 50 residues at a time throughout the molecule. The initial crystallographic R-factor using all data between 8 and 3.3 Å was 0.38 based on the stereochemically regularized structure determined from the 3.3 Å MIR map (Ollis *et al.*, 1985). The resolution of data used in the refinement was extended by periodically adding data in 0.2 Å shell increments to 2.6 Å resolution. After rebuilding and 95 cycles of refinement, the R-factor was 0.24. Forty-five water molecules were added and individual tightly restrained B-factors were refined. The computer program Geom (G.Cohn, personal communication) was used to identify residues with poor stereochemistry. Main chain residues 780−799 were retraced. In addition, 10 residues (770−779) previously built into the model as alanine were changed to their appropriate side chains and 30 residues (779−788, 617−622, 324−326, 570−575) that had not been included in the initial refinement because they were either not apparent or poorly resolved in the original MIR maps were added. The resulting R-factor was 0.198 with r.m.s. deviation in bond lengths from ideality of 0.016 Å.

Table I. Refinement summary: Klenow−dTMP complex

Stereochemistry
  r.m.s. deviation from ideality

Bond lengths (Å)	0.013
Bond angles (°)	2.7
Dihedral angles (°)	23.4
Impropers (°)	1.1

  r.m.s. deviation B-factors

Bonds (Å^2)	1.18
Angles (Å^2)	1.40

R-factor[a] = 0.185, 8−2.6 Å resolution
        = 0.196, 20−2.6 Å resolution[b]

4783	non-hydrogen atoms
87	water molecules
20 175	unique reflections used in refinement

$$^aR\text{-factor} = \Sigma_{hkl} \frac{||F_o^{hkl}| - |F_c^{hkl}||}{\Sigma_{hkl}|F_o^{hkl}|}$$

where F_o and F_c are the observed and calculated structure factor amplitudes for reflections hkl.
[b]Using bulk solvent correction.

Refinement was continued using two cycles of simulated annealing using X-PLOR (Brünger *et al.*, 1987) following the slow cooling protocol (Brünger *et al.*, 1990) and using the modified force constants reported in Weis *et al.* (1990). Because of the moderate resolution of the diffraction pattern and the large number of atoms being refined, the weight on the stereochemical terms was increased 2-fold ($W_a = 180\,000$) over the estimate initially determined by X-PLOR. The stereochemical weight was determined iteratively, by studying the r.m.s. deviation in bond length and bond angles from ideality after 50 cycles of positional refinement at an estimated weight and adjusting the value so that the deviation from ideality was 0.014 Å for bonds and 2.9° for angles. Bulk solvent correction (A.T.Brünger, L.S.Beese and W.I.Weis, unpublished) was used to calculate maps using data between 2.6 and 20 Å resolution. Inclusion of the low-resolution terms enabled 30 poorly ordered residues between 575 and 617—the 'disordered domain'— that were uninterpretable in the original MIR map to be fitted into the electron density map. Twelve out of the 604 residues of Klenow fragment have not been built into the model and are located in this region.

Residues involved in crystal lattice contacts (not handled in the PROLSQ refinement) were improved by the first cycle of X-PLOR, but regions defined by weak electron density (such as loops with B-factors >50 Å^2) needed to be restored to their starting positions. Although the structure was largely unchanged, different electron density maps that were calculated after simulated annealing refinement suggested several remaining incorrectly fitting regions. Simulated annealing 'omit' maps—electron density maps calculated after running low-temperature simulated annealing (3000°C) and positional refinement with coordinates from suspect regions omitted—aided in visualizing correct conformation. As a result the main chain between residues 750−765 and 362−367 was shifted by one residue. The structure thus did not change appreciably after the second cycle of simulated annealing. At this stage refinement had converged in the sense that residues whose positions were well determined by the electron density (generally those with B-factors <50 Å^2) returned to essentially the same positions (regions defined by weak electron density with B-factor >50 Å^2 had to be restored to starting positions). The r.m.s. deviation of main chain atoms was 0.35 Å, commensurate with the coordinate error estimated from a Luzzatti plot (Luzzatti, 1952) at this resolution. Statistics defining final stereochemistry and R-factor are described in Table I. Details of the refinement and description of secondary structure, rebuilt and added residues, and details of the polymerase active site will be described elsewhere (L.S.Beese and T.A.Steitz, in preparation).

Single-stranded DNA complex

A difference electron density map calculated using the diffraction amplitudes of the tetranucleotide complex $[F(dT_4)]$ minus those of the uncomplexed protein $[F(nat)]$ and phases from the refined Klenow model was used to position three nucleotides of the tetranucleotide complex. Although some electron density for the fourth nucleotide was visible at this stage, its

orientation was somewhat ambiguous and was therefore omitted from the initial refinement. The starting model for refinement was the native Klenow structure with Asp424 changed to an Ala, metal ion B and water molecules around the exonuclease active site omitted, and the three newly fitted nucleotides added. The initial R-factor with three of the four nucleotides fitted was 0.25 calculated between 8 and 3.1 Å resolution. Inspection of difference electron density maps indicated that there were no major changes in the protein in the polymerase domain (residues 650–928) compared with native Klenow. Because of the lower resolution of this data set compared with the native, residues in the polymerase domain (residues 670–928) were held fixed to reduce the number of terms being refined. The structure was refined using X-PLOR least-squares positional minimization. The stereochemical weight was adjusted as described for the native so that the r.m.s. deviation in bond lengths did not exceed 0.013 Å. The model of the complex was refined to convergence with an R-factor of 0.20. The fourth nucleotide, two conformations for His660, and 20 water molecules were added. Side chains in the vicinity of the tetramer were rebuilt. Positional refinement with group B-factor for the nucleotides was continued until $F_o - F_c$ difference maps showed no features above 3σ. The final R-factor was 0.185 with 0.013 Å r.m.s. deviation in bond lengths and 2.8° deviation in bond angles.

Acknowledgements

References

Atkinson,M.R., Deutscher,M.P., Kornberg,A., Russell,A.F. and Moffett,J.G. (1969) *Biochemistry*, **8**, 4897–4904.

Beese,L.S. and Steitz,T.A. (1989) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Springer-Verlag, Berlin, Vol. 3. pp. 28–43.

Bernad,A., Bianco,L., Lazaro,J.M., Martin,G. and Salas,M. (1989) *Cell*, **59**, 219–228.

Brünger,A.T., Kuriyan,J. and Karplus,M. (1987) *Science*, **235**, 458–460.

Brünger,A.T., Kurkowski,A. and Erickson,J.W. (1990) *Acta Crystallogr.*, in press.

Brutlag,D., Atkinson,M.R., Setlow,P. and Kornberg,A. (1969) *Biochem. Biophys. Res. Commun.*, **37**, 982–989.

Brutlag,D. and Kornberg,A. (1972) *J. Biol. Chem.*, **247**, 241–248.

Camerman,N., Fawcett,J.K. and Camerman,A. (1976) *J. Mol. Biol.*, **107**, 601–621.

Catalano,C.E. and Benkovic,S. (1989) *Biochemistry*, **28**, 4374–4382.

Christianson,D.W. and Lipscomb,W.N. (1988) *Colloq. Ges. Biol. Chem.*, **39**, 65–74.

Christianson,D.W. and Lipscomb,W.N. (1989) *Accts. Chem. Res.*, **22**, 62–69.

Coleman,J.E. (1987) In Torriani-Gorini,A. *et al.* (eds), *Phosphate Metabolism and Cellular Regulation in Microorganisms*. American Society for Microbiology, Washington, DC, pp. 127–138.

Cowart,M., Gibson,K.J., Allen,D.J. and Benkovic,S.J (1989) *Biochemistry*, **28**, 1975–1983.

Derbyshire,V., Freemont,P.S., Sanderson,M.R., Beese,L.S., Friedman,J.M., Steitz,T.A. and Joyce,C.M. (1988) *Science*, **240**, 199–201.

Derbyshire,V., Grindley,N.D.F., Joyce,C.M. (1991) *EMBO J.*, **10**, 17–24.

Fersht,A.R., Knill-Jones,J.W. and Tsui,W.-C. (1982) *J. Mol. Biol.*, **156**, 37–51.

Freemont,P.S., Friedman,J.M., Beese,L.S., Sanderson,M.R. and Steitz,T.A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8924–8928.

Freemont,P.S., Ollis,D.L., Steitz,T.A. and Joyce,C.M. (1986) *Proteins*, **1**, 66–73.

Gupta,A.P. and Benkovic,S.J. (1984) *Biochemistry*, **23**, 5874–5881.

Hendrickson,W.A. (1985) *Methods Enzymol.*, **115**, 252–270.

Herschlag,D. and Jencks,W.P. (1987) *J. Am. Chem. Soc.*, **109**, 4665–4674.

Holmes,M.A. and Matthews,B.W. (1982) *J. Mol. Biol.*, **160**, 623–639.

Jones,T.A. (1985) *Methods Enzymol.*, **115**, 157–171.

Joyce,C.M. (1989) *J. Biol. Chem.*, **264**, 10858–10866.

Joyce,C.M., Ollis,D.L., Rush,J., Steitz,T.A., Konigsberg,W.H. and Grindley,N.D.F. (1986) In Oxender,D. (ed.), *Protein Structure, Folding and Design*. Alan R.Liss, New York, pp. 197–205.

Joyce,C.M. and Steitz,T.A. (1987) *Trends Biochem. Sci.*, **12**, 288–292.

Kim,E.E. and Wyckoff,H.W. (1991) *J. Mol. Biol.*, in press.

Kim,H. and Lipscomb,W.N. (1990) *Biochemistry*, **29**, 5546–5555.

Konnert,J.H. and Hendrickson,W.A. (1980) *Acta Crystallogr.*, **36A**, 344–350.

Kornberg,A. (1980) *DNA Replication*. W.H.Freeman, San Francisco, CA.

Leavitt,M.C. and Ito,J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 4465–4469.

Lehman,I.R. and Richardson,C.C. (1964) *J. Biol. Chem.*, **239**, 233–241.

Liljas,A., Kannan,K.K., Bergsten,P.-C., Waara,I., Fridborg,K., Strandberg,B., Carbom,U., Järup,L., Löogren,S. and Petef,M. (1972) *Nature New Biol.*, **235**, 131–137.

Lindskog,S. and Coleman,J.E. (1973) *Proc. Natl. Acad. Sci. USA*, **70**, 2505–2508.

Luzzati,V. (1952) *Acta Crystallogr.*, **5**, 802–810.

Maki,H. and Kornberg,A. (1985) *J. Biol. Chem.*, **260**, 12987–12992.

Matthews,B.W. (1988) *Accts. Chem. Res.*, **21**, 333–340.

Ollis,D.L., Brick,P., Hamlin,R., Xuong,N.G. and Steitz,T.A. (1985) *Nature*, **313**, 762–766.

Read,R. (1986) *Acta Crystallogr.*, **A42**, 140–149.

Rould,M.A., Perona,J.J., Söll,D. and Steitz,T.A. (1989) *Science*, **246**, 1135–1142.

Scheuermann,R.H. and Echols,H. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7747–7751.

Spicer,E.K., Rush,J., Fung,C., Reha-Kranz,L.J., Daram,J.D. and Konigsberg,W.H. (1988) *J. Biol. Chem.*, **263**, 7478–7486.

Steitz,T.A., Beese,L.S., Freemont,P.S., Friedman,J.M. and Sanderson,M.R. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 465–471.

Steitz,T.A. and Lipscomb,W.N. (1965) *J. Am. Chem. Soc.*, **87**, 2488–2489.

Weis,W.I., Brunger,A.T., Skehel,J.J. and Wiley,D.C. (1990) *J. Mol. Biol.*, **212**, 737–761.

Xuong,N.H., Nielsen,C., Hamlin,R. and Anderson,D. (1985) *J. Appl. Crystallogr.*, **18**, 342–350.

Westheimer,F.H. (1968) *Acc. Chem. Res.*, **1**, 70–79.

Wyckoff,H.W. (1987) In Torriani-Gorini,A. *et al.* (eds), *Phosphate Metabolism and Cellular Regulation in Microorganisms*. American Society for Microbiology, Washington, DC, pp. 118–126.