Articles

Crystal Structures of an NH_2-Terminal Fragment of T4 DNA Polymerase and Its Complexes with Single-Stranded DNA and with Divalent Metal Ions†

J. Wang,‡ P. Yu,‡,§ T. C. Lin,‡,|| W. H. Konigsberg,‡ and T. A. Steitz*,‡,⊥,#

Department of Molecular Biophysics and Biochemistry, Department of Chemistry, and Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06520-8114

Received January 24, 1996; Revised Manuscript Received April 9, 1996⊗

ABSTRACT: We report the crystal structure of an NH_2-terminal 388-residue fragment of T4 DNA polymerase (protein N388) refined at 2.2 Å resolution. This fragment contains both the 3′−5′ exonuclease active site and part of the autologous mRNA binding site (J. D. Karam, personal communication). The structure of a complex between the apoprotein N388 and a substrate, $p(dT)_3$, has been refined at 2.5 Å resolution to a crystallographic *R*-factor of 18.7%. Two divalent metal ion cofactors, Zn(II) and Mn(II), have been located in crystals of protein N388 which had been soaked in solutions containing Zn(II), Mn(II), or both. The structure of the 3′−5′ exonuclease domain of protein N388 closely resembles the corresponding region in the Klenow fragment despite minimal sequence identity. The side chains of four carboxylate residues that serve as ligands for the two metal ions required for catalysis are located in geometrically equivalent positions in both proteins with a rms deviation of 0.87 Å. There are two main differences between the 3′−5′ exonuclease active site regions of the two proteins: (I) the OH of Tyr-497 in the Klenow fragment interacts with the scissile phosphate in the active site whereas the OH of the equivalent tyrosine (Tyr-320) in protein N388 points away from the active center; (II) different residues form of the binding pocket for the 3′-terminal bases of the substrate. In the protein N388 complex the 3′-terminal base of $p(dT)_3$ is rotated approximately 60° relative to the position that the corresponding base occupies in the $p(dT)_3$ complex with the Klenow fragment. Finally, a separate domain (residues 1−96) of protein N388 may be involved in mRNA binding that results in translational regulation of T4 DNA polymerase (Pavlov & Karam, 1994).

The DNA polymerase of bacteriophage T4 is an essential component of the T4-encoded DNA replication complex. In addition to polymerase activity it has an editing 3′−5′ exonuclease activity located in a region proximal to the polymerase domain (Spicer *et al*., 1988), a situation similar to that observed in the Klenow fragment (KF)[1] of *Escherichia coli* DNA polymerase I. However, the exonuclease activity of the T4 enzyme is about 10^3 times greater than that of KF (Huang & Lehman, 1972; Capson *et al*., 1992; Lin *et al*., 1994; Bloom *et al*., 1994).

Because the sequence identity between these two 3′−5′ exonuclease domains is so limited (15%), an alignment of part of the T4 and KF sequences (Bernad *et al*., 1989) was made that appeared to be incorrect (Morrison *et al*., 1991). Bernad *et al*. (1989) proposed that there are three short regions that can be aligned in both eukaryotic and prokaryotic editing exonucleases which they named the Exo I, II, and III motifs. Though many of the sequences aligned in the Exo motif I by Bernad *et al*. (1989) appear to be correct, some alignments including that of the T4 pol exonuclease domain do not. Reha-Krantz *et al*. (1991) mutated T4 pol residue Glu-191, found only a modest change in activity, and suggested that the 3′−5′ exonuclease mechanism of T4 polymerase might be different from KF. Morrison *et al*. (1991) concluded from mutations of yeast DNA polymerase II that the earlier sequence alignment of T4 polymerase in the Exo I region was incorrect. Including the changes in sequence alignments in Exo I, it appears that the three Exo motifs in the sequences of prokaryotic and eukaryotic polymerases represent conserved localized regions found in the polymerases from bacteriophage T4 and other B family DNA polymerases that have editing exonuclease functions. The four essential carboxylates that serve as ligands for the two divalent metal ions are found in these motifs, suggesting that all of these polymerases employ the same two-metal-ion catalytic mechanism found for KF (Freemont *et al*., 1988; Beese & Steitz, 1991). While the exonuclease domains of the 388-residue N-terminal fragment of T4 DNA polymerase (protein N388) and KF are structurally related, the functional differences between the two enzymes require explanation.

† This research was supported by American Cancer Society Grant BE-52J to T.A.S. and USPHS Grant GM12607-31 to W.H.K.

* Corresponding author.
‡ Department of Molecular Biophysics and Biochemistry.
§ Present address: Bristol-Myers Squibb, Syracuse, NY 13220.
|| Present address: MicroGeneSys, Meriden, CT 06450.
⊥ Department of Chemistry.
Howard Hughes Medical Institute.

⊗ Abstract published in *Advance ACS Abstracts,* June 1, 1996.

[1] Abbreviations: KF, Klenow fragment of *Escherichia coli* polymerase I; SIR/AS, single isomorphous replacement with anomalous scattering; EMP, ethylmercury phosphate; protein N388, a fragment of T4 DNA polymerase consisting of residues 1−388.

S0006-2960(96)00178-X CCC: $12.00

Table 1: X-ray Diffraction Data

space group: $P2_12_12$
unit cell: $a = 114.69$ Å, $b = 109.84$ Å, $c = 70.28$ Å

data sets	resolution (Å)	$R_{merging}$ (%)	footnote
native I	3.2	8.1	*a*
native II	2.7	8.0	*a*
EMP	2.9	9.3	*a*
Zn	3.5	11.5	*b*
Zn, Mn	2.8	4.0	*b*
Zn, Mn, Mg	3.0	6.3	*b*
$p(dT)_3$ I	3.1	7.4	*c*
$p(dT)_3$ II	2.5	13.9	*c*
native III	2.27	4.0	*d*
native IV	2.20	9.6	*e*

[a] All three data sets of native I and native II and ethylmercury phosphate (EMP), soaked at 1 mM for 12 h, were collected using Hamlin's multiwire area detector system at room temperature. [b] All divalent metal ion soaking data sets were collected using Hamlin's detector system at room temperature. Concentrations of Zn(II), Mn(II), and Mg(II) were 0.5, 5, and 5 mM, respectively. Divalent metal soaked at other concentrations (not listed in this table) had similar quality. Concentrations of each of the divalent metal ions were varied, and their real and anomalous occupancies were refined using the heavy atom parameter refinement program MLPHARE phased by EMP single isomorphous replacement with anomalous scattering (SIR/AS). [c] Lower resolution $p(dT)_3$ data set was collected using Hamlin's detector system, and higher resolution data set was collected using the MacScience imaging-plate system, both at room temperature. Lower resolution data set and native data sets were isomorphous with each other. [d] Native III data were collected on an R-axis imaging-plate system at 110 K with unit cells of $a = 113.62$ Å, $b = 109.31$ Å, and $c = 68.58$ Å. [e] Native IV data were collected on MacScience at room temperature. This data set and the other two native data sets (native I and native II) were isomorphous with each other but not with native III.

The intact 898-residue T4 DNA polymerase, unlike KF, is able to regulate translation of its own messenger RNA by binding to a specific RNA sequence just 5′ to its translation initiation signal (Pavlov & Karam, 1994). Even though protein N388 is severely truncated, it still retains the ability to bind specifically to its own mRNA although with lower affinity than the intact enzyme (J. D. Karam, personal communication). Further truncation of protein N388, by removal of 95 residues from its N terminus had little effect on its specific exonuclease activity, indicating that the active center for phosphodiester bond hydrolysis is located between residues 96 and 388 (Lin *et al*., 1994).

In this report we describe the crystal structure of protein N388 as well as the structures of two separate complexes, one formed with divalent metal ions and the other with a short DNA substrate, $p(dT)_3$. The structure shows that the sequence alignment between the T4 and KF 3′−5′ exonuclease domains proposed by Morrison *et al*. (1991) is indeed correct, and contrary to earlier concerns (Reha-Krantz *et al*., 1991), the exonuclease activity of T4 DNA polymerase appears to employ the same two-metal-ion enzymatic mechanism as KF.

EXPERIMENTAL PROCEDURES

Crystals and Data Collection. Protein N388 derived from T4 DNA polymerase was obtained from a high-level expression vector containing the coding region for residues 1−388 of T4 DNA polymerase as described previously (Lin *et al*., 1994). The protein crystallized at 4 °C in 2.0 M sodium formate buffered with 5 mM Tris-HCl at pH 7.0 in the presence of 0.1 mM EDTA, a condition identified by screening the 50 conditions of the sparse matrix crystallization kit (Jancarik & Kim, 1991). Crystals belong to the orthorhombic space group $P2_12_12$, with unit cell dimensions of 114.7 Å × 109.8 Å × 70.3 Å. They contain two molecules per asymmetric unit (molecules 1 and 2) resulting in a V_m of 2.5 Å^3/Da and 50% solvent content (Matthews, 1968).

Although most crystals diffracted to only 3.2 Å resolution at room temperature, one 2.2 Å resolution native data set and one 2.5 Å resolution data set of protein N388 complexed with $p(dT)_3$ were collected on a dip 2000 imaging-plate system made by MacScience. On average, 1 in every 15 crystals diffracted beyond 2.4 Å. Intensities were integrated using the program DCREDUCE (Xuong *et al*., 1985a) for the data measured using the San Diego multiwire system (Xuong *et al*., 1985b) and the program DENZO (Z. Otwinowski and W. Minor, in preparation) for data obtained from the imaging-plate system. The integrated intensities were then scaled using the program SCALEPACK (Otwinowski and Minor, in preparation). A summary of the data collection is shown in Table 1.

Heavy Atom Derivatives and Structure Determination. Initial phases were determined using single isomorphous replacement and anomalous scattering (SIR/AS) differences from an ethylmercury phosphate derivative (EMP) prepared by soaking a crystal in solutions containing 2.0 M sodium formate and 1 mM EMP. Both native and derivative data are in the Brookhaven Protein Data Bank with accession numbers R1NOYSF and R1NOZSF, respectively. A difference Patterson map calculated at 3.2 Å resolution using the native data set (native I of Table 1) contained peaks of heights greater than 16σ from which two mercury sites were located. Heavy atom parameters were refined, and protein phases were calculated with the program MLPHARE (Ot-

Table 2: SIR/AS Heavy Atom Refinement

phasing EMP−native I									
resolution (Å)	22.7	12.0	8.2	6.2	5.0	4.2	3.6	3.2	total
phasing power	8.29	5.74	5.26	4.80	4.12	2.58	1.68	1.10	2.18
no. of reflections	34	232	588	1126	1882	2776	3451	4274	14363
figure of merit	0.88	0.76	0.85	0.74	0.78	0.67	0.48	0.30	0.545
phasing EMP−native II									
resolution (Å)	20.4	10.8	7.3	5.5	4.4	3.7	3.2	2.8	total
phasing power	3.08	6.13	4.97	5.21	3.23	2.19	1.67	1.02	2.15
no. of reflections	64	393	930	1717	2683	3724	4806	5967	20284
figure of merit	0.67	0.89	0.86	0.86	0.69	0.55	0.45	0.29	0.517
heavy atom	X	Y	Z	Q	Qanom	B			
Hg(I)	0.16184	0.10485	0.53213	63.79	5.89	28.60			
Hg(II)	0.74976	0.11081	−0.0033	69.45	6.75	56.55			

winowski, 1991). The overall phasing power was 2.2, and the figure of merit was 0.55 (Table 2). SIR/AS phases were improved using the program SQUASH (Zhang, 1993). After density modification, the map correlation changed overall by 35%, and the average phase angle change was 37°. Electron density maps were displayed and interpreted on ESV and SGI graphics work stations with the program O (Jones *et al*., 1991). At this stage, 85% of the amino acid residues could be unambiguously assigned.

Twofold noncrystallographic symmetry (NCS) averaging of the electron density map was employed at an early stage using the RAVE package (Kleywegt & Jones, 1994) in conjunction with the CCP4 package (SERC, 1979) for Fourier transformation and for phase combination and the program O for mask editing and displaying. A map correlation coefficient between the electron density of the two molecules was initially 64% and converged at 90% after 20 cycles of averaging; the *R*-factor between amplitudes calculated from the averaged density and observed amplitudes was 17.4% Although NCS averaging helped to interpret molecule 2, the map averaging did not yield the best map, since the two copies of the molecule had different thermal motions in the lattice.

After the structure was initially refined, we were able to further improve the experimental phases using a better native data set (native II of Table 1). Using the native II data set the mercury peaks in the isomorphous 3.2 Å resolution difference Patterson map were 17σ, the overall phasing power was 2.2 for all data to 2.8 Å resolution, and the mean figure of merit was 0.52. The phases were further changed by 42° after density modification using the program SQUASH (Zhang, 1993), resulting in an overall map change of 33% in the map correlation coefficient. In this map, many carbonyl groups are clearly visible, and nearly the entire molecule 1 is visible at a 2σ contour level, with the exception of residues 45−50 and 249−256, which were visible at 1σ. A portion of this electron density is shown in Figure 1. Molecule 2, on the other hand, had a relatively poor quality of electron density and could be traced independently only at density levels of 1σ, with the exception of residues 44−51, 73−81, and 250−255, which were not visible, and discontinuous densities around residues 228−234 and 296−298.

The initial model, consisting of 368 residues per molecule, yielded a crystallographic *R*-factor of 47.8% for data between 15 and 3.2 Å resolution and fell to 31.5% after Cα restrained positional refinement using X-PLOR (Brünger, 1992). Simulated annealing of the model using all data with $F \geq 2\sigma$ between 15 and 2.9 Å resolution further decreased the *R*-factor to 24.9%. Finally, the coordinates were refined against native data set III, resulting in a crystallographic *R*-factor of 22.2% for all data between 6 and 2.2 Å resolution (Table 3). Thermal motion differences between two independent molecules in the asymmetric unit seem to be smaller with native data set III than with earlier data sets. A portion of the final electron density map calculated using $2F_o - F_c$ amplitudes and phases from the refined structure (pdb accession number 1NOY) is shown in Figure 1. Currently, residues 2−373 are included in the structure of molecule 1, but molecule 2 is missing residues 43−51, 73−81, and 250−255. We have also refined at 2.3 Å resolution the apo structure determined at 150 K (pdb accession number 1NOZ).

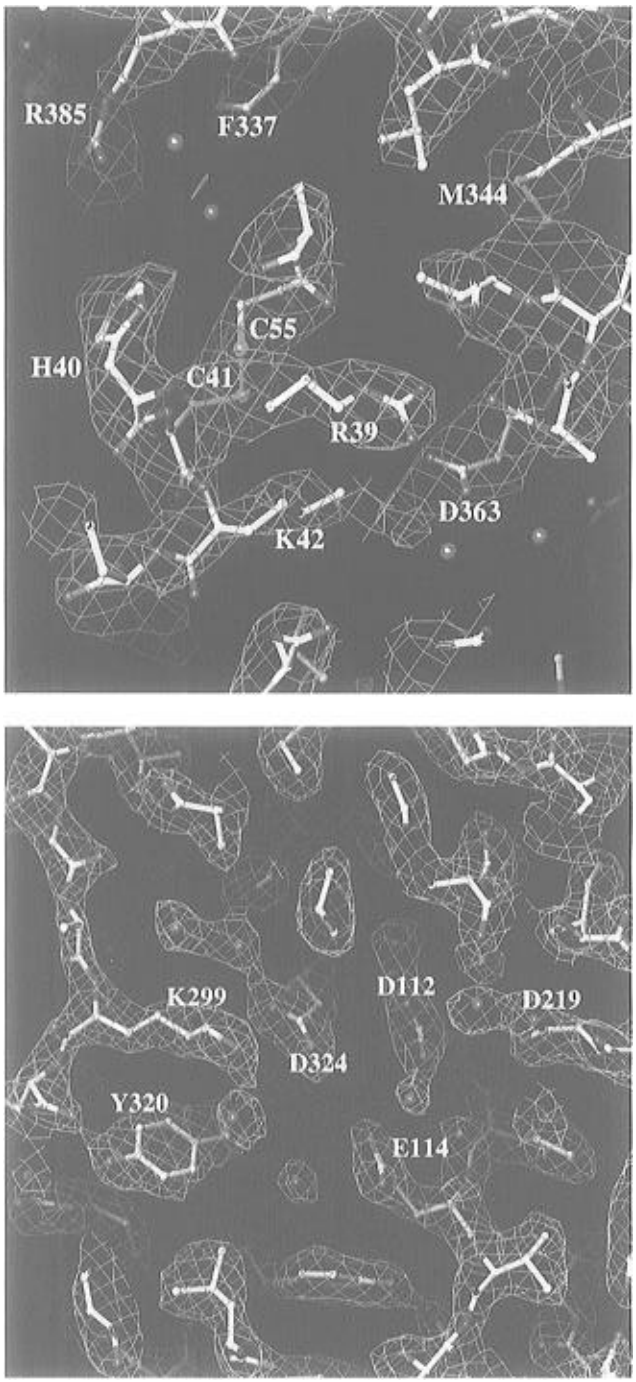


FIGURE 1: Electron density maps. (a, top) An experimental electron density map which was calculated using SIR/AS phasing and improved with the program SQUASH. The map is superimposed on a refined model with an anomalous difference Fourier map superimposed which shows the presence of a disulfide bond (Cys-41/Cys-55) that is unexpected in this cytoplasmic protein. All 22 sulfur atoms in molecule 1 and about 14 of 22 in molecule 2 are unarbitrarily located in an experimental native anomalous difference Fourier map. (b, bottom) The $2F_o - F_c$ electron density map calculated from the refined structure showing an unexpected orientation of Tyr-320 and a unique positively charged residue (Lys-299) found in many B family polymerases. It is noted that Lys-299 interacts in a slightly different manner for apo and the apo−$p(dT)_3$ complex (see Figure 5).

However, only one-quarter of the ordered solvent molecules of 1NOY at room temperature were observed in 1NOZ at the liquid N_2 temperature.

Comparison of the 3′−5′ Exonuclease Domain Structures. Comparison of the structure of the 3′−5′ exonuclease domain

Table 3: Structure Refinement

no. of non-hydrogen atoms	5896
no. of solvent molecules	179
rms bond deviation (Å)	0.010
rms angle deviation (deg)	1.54
rms dihedral angle deviation (deg)	24.40
rms improper angle deviation (deg)	1.4
resolution limits (Å)	6.0−2.2
no. of reflections (2σ cutoff)	37292
Wilson *B*-factor (Å^2)	39.8
crystallographic *R*-factor (%)	22.2
estimated coordinate error (Å)	0.3
average *B*-factor (mc, *mol* 1) (Å^2)	40.8
average *B*-factor (mc, mol 2) (Å^2)	42.3
average *B*-factor (sc,[a] mol 1) (Å^2)	55.0
averge *B*-factor (sc, mol 2) (Å^2)	55.9
rms ΔB (mc, 1−2 bonded)[a] (Å^2)	1.6
rms ΔB (mc, 1−3 angle)[a] (Å^2)	2.2
rms ΔB (sc, 1−2 bonded) (Å^2)	2.9
rms ΔB (sc, 1−3 angle) (Å^2)	3.6

[a] mc and sc stand for main chain and side chain atoms, respectively. 1−2 bonded and 1−3 angle bonded ΔB are temperature factor differences between directly bonded atoms and indirectly bonded atoms via a shared atom, respectively.

protein N388 with that of KF was carried out by a stucture-based sequence alignment in the program O as shown in Figure 3 (Jones *et al*., 1991).

RESULTS

Structure of Protein N388. The structure of protein N388 consists of two domains (Figure 2) as predicted from the results of limited proteolysis (Lin *et al*., 1994). The larger domain (residues 107−338) contains the exonuclease active site and has a folding pattern nearly identical to that of KF (Ollis *et al*., 1985). The smaller domain, comprising residues 1−106, is the region that contributes to autogenous mRNA recognition. The main structural feature of the larger domain is a five-stranded β sheet formed by strands numbered 6−10 and by six α-helices. The last 35 residues (355−370) at the carboxyl terminus, exiting from the exonuclease domain, are folded back onto the smaller domain in this fragment. The conserved Exo I, Exo II, and Exo III sequence motifs, which have been found in all DNA polymerases with 3′−5′ exonuclease activity (Morrison *et al*., 1991; Blanco *et al*., 1992), are located in β-6, α-D, and α-H, respectively (Figure 2). An additional motif, Exo I′, is located in a loop connecting β-9 to α-C in the back of the exonuclease active site (Figure 2). This motif is highly conserved among B family DNA polymerases but absent in the pol I (A family). Because the Exo I′ motif in the polymerase B family closely resembles the Exo I motif of the pol I family, they were previously misaligned (Spicer *et al*., 1988).

The main structural features of the smaller domain are two antiparallel β sheets and two α-helices (A and B). Additionally, three α-helices (I, J, K) from the C terminus of this fragment are packed against the small NH_2-terminal domain. The larger β sheet is a β meander formed by stands 1, 2, and 3 while the small one is a β hairpin formed by strands 4 and 5. The structure of this domain provides no obvious insights at the moment into how it might participate in the binding of mRNA.

Similarity between the 3′−5′ Exonuclease Domains of KF and Protein N388. Comparison of the structure of protein N388 with that of the 3′−5′ exonuclease domain of KF shows extensive similarity at the metal binding sites and diminishing similarity away from the active site. The root mean square (rms) difference between 110 equivalent Cα coordinates (Figure 3), which include conserved motifs Exo, I, II, III, and I′, is 1.8 Å. There are two large insertions (each more than 30 residues) in protein N388 compared to KF, one after residue 149 (β-8) and the other after residue 240 (α-D); extrapolating from yeast pol I data and sequence alignments of region IV (Wang *et al*., 1989) the former insertion may be near a putative primase interaction region, a function that is not required of *E. coli* pol I.

There are six highly conserved residues within the three Exo motifs. Conserved Asp-112 and Glu-114 in the Exo I motif of protein N388 are equivalent to Asp-355 and Asp-357 in KF. Within the Exo II and Exo III motifs there are two additional conserved aspartate residues, Asp-219 and Asp-324 in protein N388 which correspond to Asp-424 and Asp-501 in KF. The side chains of these four acidic residues from protein N388 and KF can be superimposed with a rms difference of 0.87 Å. As discussed in the following section, these carboxylate groups serve as ligands for the two metal ions required for catalysis in proton N388, as found earlier for KF. The other conserved residues are Asn-214 in protein N388 which corresponds to Asn-420 in KF (Exo II motif) and Tyr-320 in protein N388 which is equivalent to Tyr-497 in (Exo III motif). In addition to the conserved sequences in the Exo II and Exo III motifs, there is either a Phe (Phe-218 in protein N388) or a Tyr (Tyr-423 in KF) in nearly all polymerase 3′−5′ exonuclease domains in a position just proximal to the conserved aspartic acid residue in each enzyme, Asp-219 in protein N388 and Asp-424 in KF (Blanco *et al*., 1992).

Divalent Metal Ions Located in the Exonuclease Active Site of Protein N388. As in the case of KF there are two metal ion binding sites (A and B) in the exonuclease active site of T4 protein N388 (Figure 4). The locations of the two metal ion sites were found by diffusing 1 mM $ZnCl_2$ and 10 mM $MnCl_2$ into the crystal, measuring data to 3.3 Å resolution, and calculating both isomorphous and anomalous difference Fourier syntheses. Site A is located almost equidistant (2.0 Å) from the carboxylates of Asp-112, Glu-114, and Asp-324 (Figure 4). The shape of the highest peak (17σ) in an isomorphous difference Fourier was elongated toward the second metal site (B) located between Asp-112 and Asp-219 in molecule 1. The anomalous difference Fourier indicates the presence of two separated sites. In molecule 2 the difference electron density map more clearly shows two peaks, one at site A and the other at site B. When only Zn(II) was diffused into the crystals, the peaks in the difference Fourier syntheses were located primarily at the metal A site, with weaker density at site B. The opposite result was obtained when only Mn(II) was soaked in: the density at site B was stronger than at site A. The distance between the two metal ions are 3.13 and 3.59 Å in molecules 1 and 2, respectively.

Crystals were soaked in various concentrations of Zn(II) and Mn(II), either metal ion alone or in combinations of both, and the occupancies of the metal ions bound to sites A and

B were refined with the heavy atom parameter refinement program MLPHARE using experimental phases. This allowed estimation of the relative affinities of sites A and B for these two metal ions under the buffer condition used for crystallization (2.0 M sodium formate) which doubtless differs from their affinities under physiological conditions. The dissociation constant for Zn(II) at site A is about 0.5 mM and that of Mn(II) at site B is about 10 mM. Further, site B has a low affinity for Zn(II), and site A shows weak affinity for Mn(II). The binding preference of Zn(II) for site A relative to site B is about 5−7-fold, while the binding preference of Mn(II) to site B relative to site A is about 3−5-fold. As noted previously for KF (Beese & Steitz, 1991), this divalent metal ion specificity is consistent with the tetrahedral coordination of site A and the octahedral coordination of site B and the coordination preferences of Zn(II) and Mn(II).

From refinement of the occupancies of the fully substituted sites A and B it is possible to conclude with assurance that there are two simultaneously occupied binding sites and not two mutually exclusive alternative binding sites. If an approximate absolute scale is established by assuming that the −SH binding sites of the ethylmercury derivative are fully occupied at 80 electrons, then the A site containing Zn(II) at 1 mM refines to 19 electrons and the B site Mn(II) at 10 mM refines to 12 electrons in two separate experiments. Finally, site A titrates to full occupancy at lower concentrations of Zn(II) than site B, consistent with two independent binding sites. When both metal ions are present at 0.5 mM Zn(II) and 10 mM Mn(II), the occupancies refine to 19 electrons at site A and 16 electrons at site B.

Structure of the Complex with a Trinucleotide Substrate. The structure of protein N388 complexed with $p(dT)_3$ has been refined at 2.5 Å resolution to a crystallographic *R*-factor of 18.7%. A difference electron density map between protein N388 and the apoprotein complexed with $p(dT)_3$ (Table 1) is shown in Figure 4. Also, superimposed in this figure are difference Fourier maps from separate experiments, described above, that show the positions of both zinc and manganese ions relative to the oligonucleotide substrate. A nonbridging oxygen from the scissile phosphate provides the fourth ligand for the metal ion at site A along with the Asp-112/Glu-114/Asp-324 carboxylates giving a tetrahedrally coordinated complex. The same nonbridging oxygen as well as the 3′-bridging oxygen from the phosphate and carboxylate oxygens of Asp-324 and Asp-219 forms part of the octahedral coordination for the metal ion at site B. Due to a blockage of the binding site in molecule 1 by a crystal contact, $p(dT)_3$ is only bound to molecule 2.

Comparison of the trinucleotide substrates bound to KF and to T4 protein N388, after superposition of their corresponding Cα backbone structures, shows that the major difference is the conformation of the 3′- and 5′-terminal bases. The 3′ base is rotated by approximately 60° and the 5′ base by about 30° when $p(dT)_3$ is bound to protein N388 relative to its conformation when bound to KF (Figure 4). The first two bases from the 3′ end in protein N388 are more nearly parallel yet too separated for direct stacking interactions on each other. Detailed protein−substrate interactions from the refined structure are shown in Figure 5. A conserved aromatic ring (Phe at position 218 in protein N388) forms extensive van der Waals contacts with the 3′-deoxyribosyl moiety and also interacts with the adjacent sugar. Phe-120 and Ile-306 in protein N388 are in similar locations as Leu-361 and Phe-473 in KF, relative to the 3′-terminal and penultimate bases, although the details of their interactions with these bases differ.

DISCUSSION

DNA polymerases are multifunctional enzymes that catalyze phosphoryl transfer reactions. In the synthetic mode, nucleotide residues are added in a template-directed fashion to the 3′ end of a growing primer. In the degradative mode, nucleotides are excised from either the 5′ and 3′ termini. These functions can reside either in the same polypeptide chain or in separate subunits in the holoenzyme (Joyce &

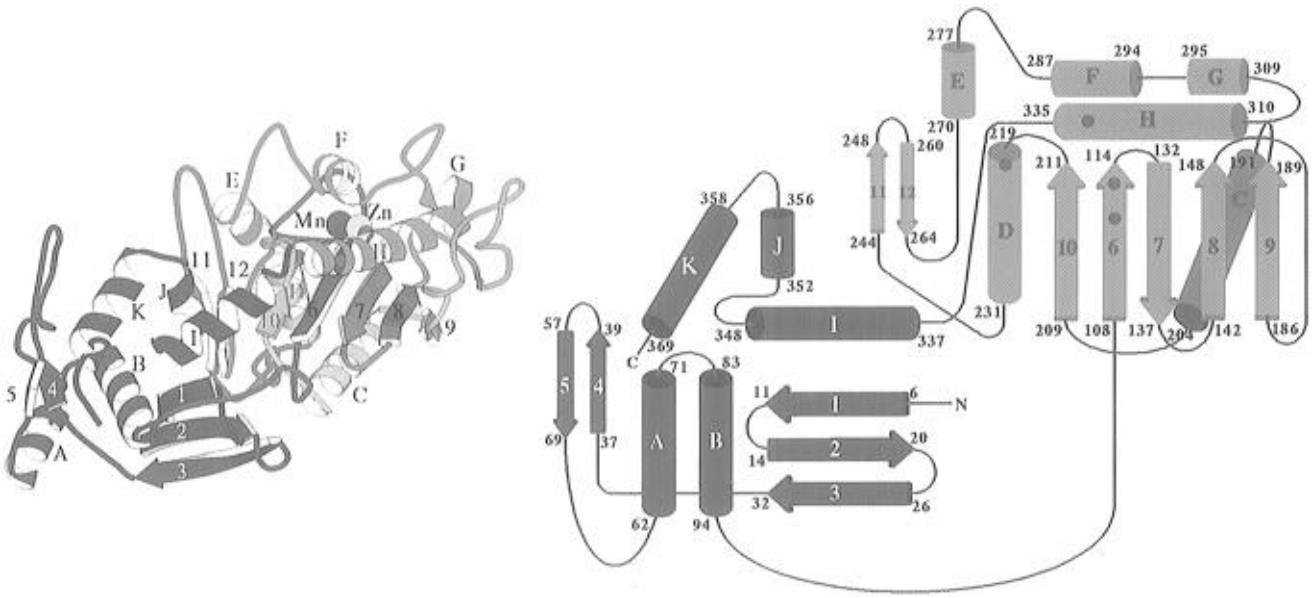


FIGURE 2: Overall structure of protein N388. (a, left) A ribbon and arrow (MOLSCRIPT) (Kraulis, 1991) drawing the overall backbone structure which is rainbow color coded from the N to C terminus. (b, right) A schematic drawing of the protein's secondary structure. The 3′−5′ exonuclease domain containing the four essential metal liganding carboxylates (dots) is shown in blue while the N-domain that may be involved in RNA binding is in red.

```
T4.  107 RVANCDIEVT 116  ...  120 FPDPMK 125  ...  129 EIDAITHY 136
KF.  350 PVFAFDVESD 359  ...  360 SLDNIS 365  ...  367 NLVGLSFA 374

T4.  145 VFDL 148  ...  188 FDNERDMLMEYI 199  ...  209 IFTGWN 214
KF.  382 YIPV 385  ...  395 QISRERALELLK 406  ...  415 LKVGQN 420

T4.  214 NIEGFDVPYIMNRV 227  ...  267 ILDYLDLYKKF 277 ...
KF.  420 N.LKYDRGILANYG 432  ...  439 AFDIMLESYIL 449 ...

T4.  284 SFSLESVAQHEIKKGK 299  ...  313 NHQRYISYNIIDVESVQAIDKIRG 336
KF.  455 RHDMDSLAERWLKHKT 470  ...  490 ALEEAGRYAAEDADVTLQLHLKMW 513
```

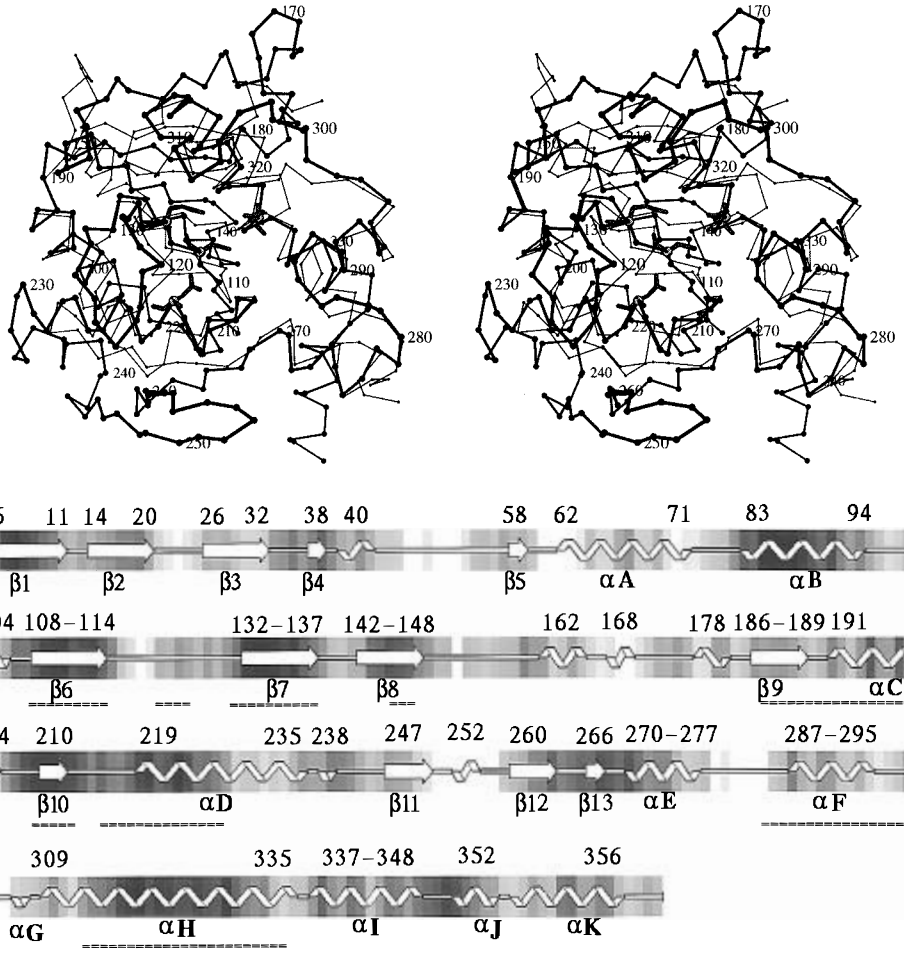
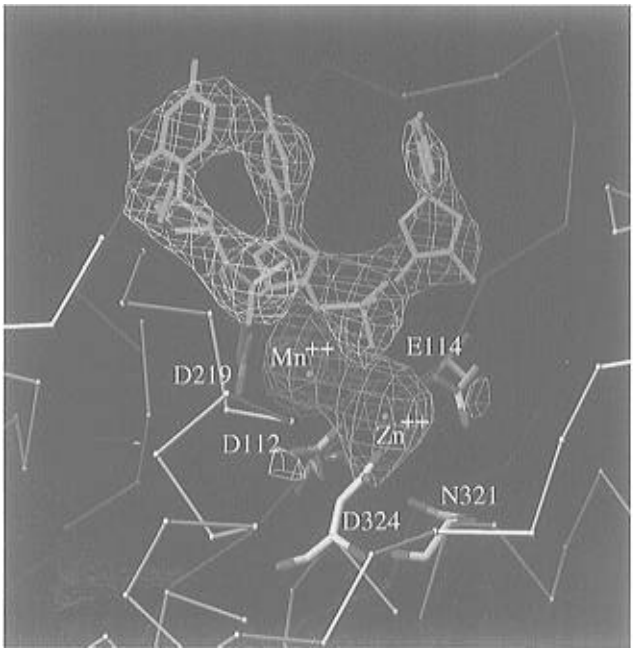
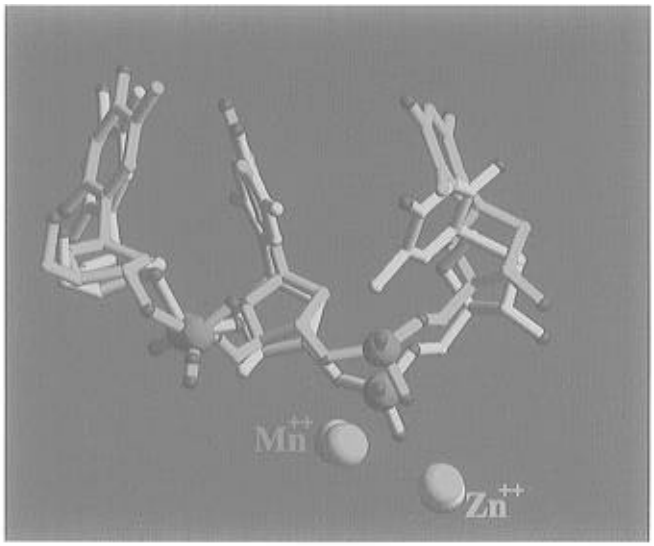


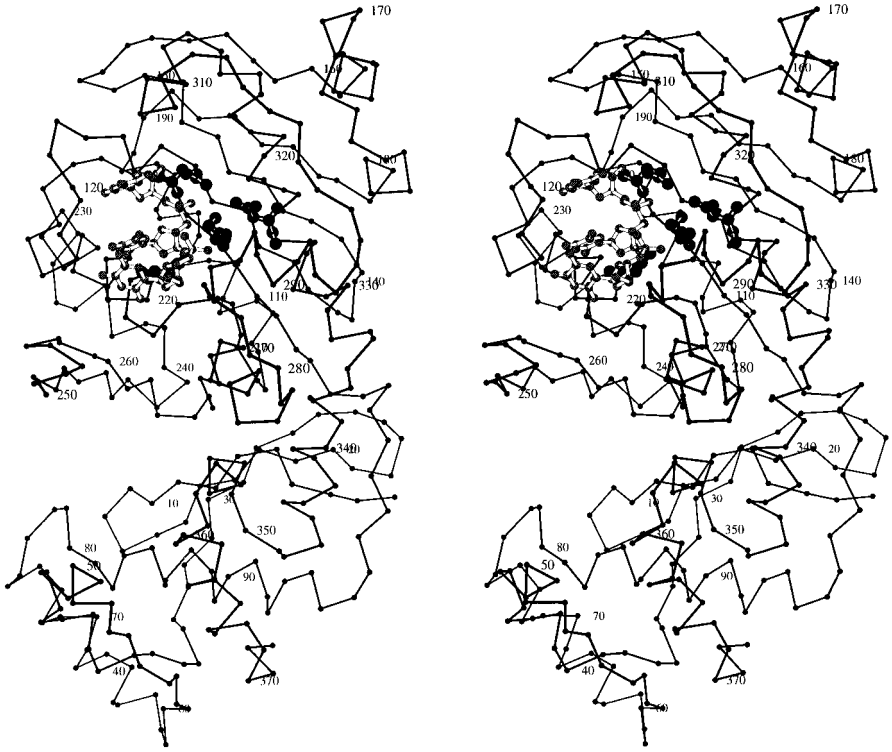
FIGURE 3: Structural comparison of the 3′−5′ exonuclease domain of protein N388 with that of the Klenow fragment. (a, top) A sequence alignment between protein N388 and KF as found from a least squares superposition of the two structures using the program O. Four conserved carboxylates are underlined. (b, middle) Stereo diagram of the Cα backbone structure of the T4 exonuclease (thick lines) superimposed on the KF (thin line) 3′−5′ exonuclease domain with every 10th residue labeled on the T4 enzyme. (c, bottom) The degree and location of sequence similarity among aligned 3′−5′ exonuclease domains between the T4 exonuclease and KF. Locations of the aligned sequences are underlined. Temperature factors in the T4 structure from low to high are indicated by shading from dark to light. The regions that are structurally conserved between T4 and KF form the core of the exonuclease domain, which has the lowest temperature factors.

Steitz, 1994). The mechanism of template-directed DNA synthesis has been derived from a combination of biochemical, genetic, and crystallographic studies of *E. coli* DNA polymerase I (Joyce & Steitz, 1994), *Thermus aquaticus* DNA polymerase (Kim *et al*., 1995), and rat DNA polymerase β (Davies *et al*., 1994; Sawaya *et al*., 1994). The catalytic activities manifested by this class of enzymes have been ascribed to structurally and functionally independent domains—an N-terminal domain with 5′ exonuclease activity and a central and C-terminal domain which are responsible for proofreading and polymerase functions, respectively. This structural organization has been extrapolated to other DNA polymerases whose primary structures are closely related to the pol I prototype. Among them, an α-like DNA polymerase has been isolated from *Sulfolobus solfataricus*, which also has its catalytic activities organized in a modular fashion (Pisani & Rossi, 1994).

The crystal structure of N388 protein now clearly confirms the modified amino acid sequence alignment of the Exo I motif (Morrison *et al*., 1991; Blanco *et al*., 1992) and supports the proposal that all DNA polymerases containing an editing 3′−5′ exonuclease have a domain whose structure and mechanism of action are homologous to that of KF. Mutation of Asp-219 to Ala (Frey *et al*., 1992; Sattar *et al*.,


D219
Mn++
E114
Zn++
D112
N321
D324


Mn++
Zn++



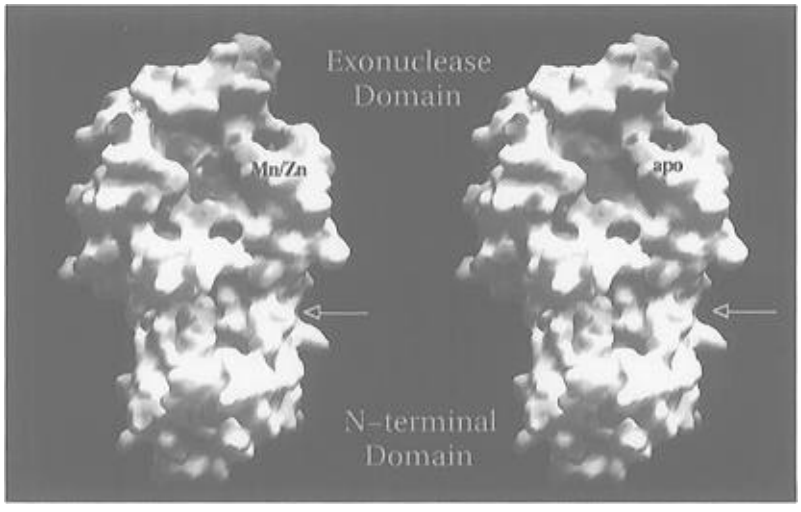

Exonuclease
Domain
Mn/Zn
apo
N-terminal
Domain

FIGURE 4: (Left page) Difference Fourier maps showing the positions of bound divalent metal ions and substrate, $p(dT)_3$. (a, top left) Superposition of a difference Fourier map between a trinucleotide complex and the apoenzyme showing the binding of the substrate $p(dT)_3$ in molecule 2, along with isomorphous and anomalous difference Fourier maps, showing the positions of bound zinc and manganese ions which are separated by 3.6 Å. In red is the anomalous difference density arising from Mn(II) bound to site B, and in light blue is the isomorphous difference density showing Zn(II) bound primarily at site A and weakly at site B. (b, top right) Superposition of $p(dT)_3$ bound to KF in yellow and $p(dT)_3$ bound to protein N388 in blue. The positions of bound divalent metal ions are shown with the same color code. Superposition was achieved by overlaying corresponding parts of the two proteins. (c, middle) Stereo α-carbon backbone drawing of protein N388 with the trinucleotide substrate and conserved carboxylates shown. (d, bottom) Solvent-accessible surface representation of protein N388 showing the active site clefts and the negative (red) and positive (blue) electrostatic charge potential both with bound Mn/Zn (left) and without (right). As with most of the nucleases using the two metal ion mechanism, the active site cleft has a preponderance of negative charges. Arrows indicate a cleft formed between the exonuclease domain and the N-terminal domain. The cleft continues into the exonuclease active site.

1996) and Asp-324 to Gly (Reha-Krantz *et al*., 1991) as well as mutations of the other two metal ion ligands, Asp-112 and Glu-114 (Nonay & Reha-Krantz, 1993; Sattar *et al*., 1996), results in a major loss of 3′−5′ exonuclease but not polymerase activity. Similar studies have also been carried out with the exonuclease domain of phage *ϕ29* where residues Asp-12, Glu-14, Asp-66, Asp-169, and Tyr-165 have been altered with concomitant loss of exonuclease activity (Esteban *et al*., 1994).

Protein N388 is clearly an independent functional domain of the T4 DNA polymerase since it has almost the same 3′−5′ exonuclease specific activity on $p(dT)_3$ as the intact T4 DNA polymerase (Lin *et al*., 1995) and also has the similar RNA binding specificity (J. D. Karam, personal communication). However, its relationship to the pol domain in the native enzyme differs from that found in the pol I (A family) since single point mutations in exo or pol domains of the B family DNA polymerases have been found that affect the activities of both the exo and pol domains (Gibbs *et al*., 1991; Soengas *et al*., 1992; Sattar *et al*., 1996). In a study, conducted in parallel with this work, several mutations affecting exonuclease activity were found between residues 96 and 359 which correspond to the larger domain of protein N388. Some of these mutations affect polymerase activity even though they are in the exo domain. The converse is also true; mutations in the region distal to residue 388 (toward the C terminus) show diminished exo activity, particularly with longer DNA substrates such as $p(dT)_{16}$ (Sattar *et al*., 1996). These studies suggest an interdependence between the polymerase and the exonuclease domains. Several mutations in the intact T4 DNA polymerase that presumably affect RNA binding (Hughes *et al*., 1987; Hsu, 1992) are located in the region spanning residues 1−96 which corresponds to the smaller N-terminal domain of protein N388. We conclude that the smaller N-terminal domain of protein N388 is part of the RNA binding region on the basis of the behavior of single site mutants from T4 DNA polymerase (Hsu, 1992) and on the failure of a proteolytic fragment of protein N388 containing residues 96−331 to bind its target RNA (W. H. Konigsberg, unpublished results). Nevertheless, we cannot rule out an additional role of the N-terminal domain in the binding of DNA substrates for the polymerase reaction, since competition experiments indicate that the protein determinants for RNA and DNA recognition partially overlap (Pavlov & Karam, 1994).

The highly conserved Tyr-320, located in the Exo motif III in protein N388, is hydrogen bonded to the backbone carbonyl of Leu-300, which is a significantly different interaction from that of the equivalent residue (Tyr-497) in KF where the OH group points toward the active site and is observed to interact with the scissile phosphate and a metal-bound water molecule (Freemont *et al*., 1988; Beese & Steitz, 1991). Although no significant structural changes were observed in a preliminary crystallographic characterization of protein N388 with Tyr-320 → Phe and Tyr-320 → Ala substitutions, biochemical studies have shown that these substitutions in the intact enzyme substantially reduce both the polymerase and exonuclease activity (Sattar *et al*., 1996). These results support the notion that the exonuclease and polymerase domains are interdependent in T4 and possibly in all other B family DNA polymerases as well. These data also imply that the OH group of Tyr-320 may be making the same substrate interactions as Tyr-497 of KF when metal ion A as well as substrate is present. There is also the likely possibility that the orientation of Tyr-320 in protein N388 is different from the conformation that it assumes in the intact enzyme. Finally, the partially conserved positively charged residue (Spicer *et al*., 1988) at position Lys-299 appears to form a salt bridge with the conserved Asp-324 carboxyl group (Figures 1 and 5). As there is no equivalent salt bridge in KF, this represents an important structural difference between the two enzymes which may have a bearing on their relative 3′−5′ exonuclease activities.

The trinucleotide backbone's overall conformation and interactions with protein N388 are largely the same as the tetranucleotide bound to KF. Benkovic and colleagues (Cowart *et al*., 1989) found that the 3′−5′ exonuclease of the T4 polymerase could digest to within two nucleotides of an interstrand cross-link, whereas the KF could digest to only four nucleotides of a cross-link, implying a difference in the size of the single-stranded DNA binding site. This structure plus the recently determined structure (J. Wang, A. Sattar, J. Karam, W. Konigsberg, and T. Steitz, unpublished) of the native DNA polymerase of phage RB69, a homologue of T4, requires that an alternative explanation be found to the different digestion of cross-linked DNA exhibited by KF and gene 43 protein.

In conclusion, the structures of the N-terminal half of T4 DNA polymerase and its complexes with metal ions and short deoxyoligonucleotides show many striking similarities to those of the 3′−5′ exonuclease domain of KF (in spite of low sequence similarity) as well as some significant differences. The mode of binding of two divalent metal ions by four protein carboxylates and the relationship of the metal ions to the scissile phosphate of single-stranded DNA appear largely the same in the two enzymes. However, the protein side chains binding the 3′-terminal two bases are different in the two exonucleases. Further, there is a partially conserved positively charged residue (Lys-299) and an

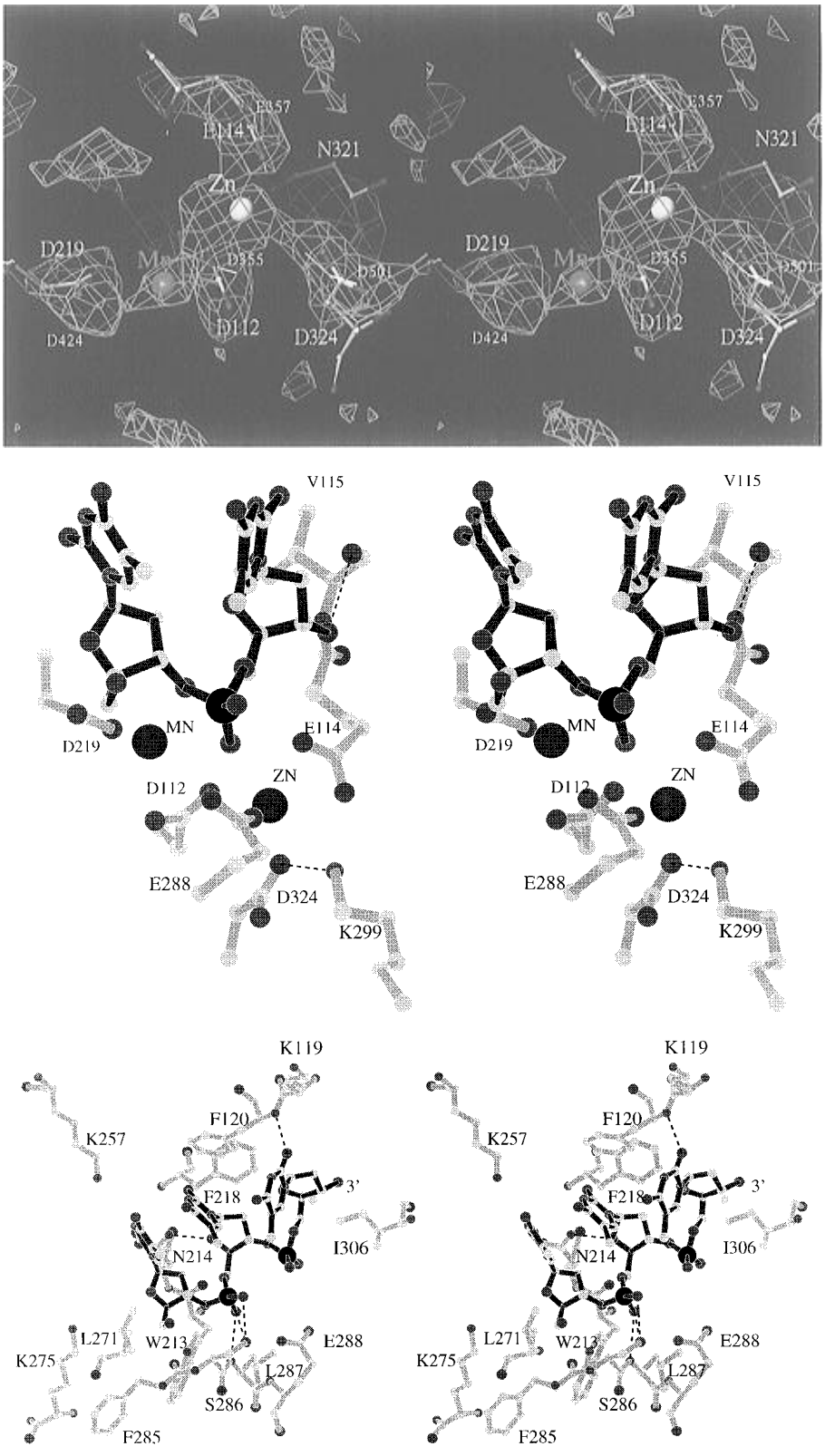


FIGURE 5: Detailed interactions between the protein, the metal ions, and the DNA substrate. (a, top) An electron density map calculated using $F_o - F_c$ as coefficients, where F_o is the observed amplitude from the metal ion complex and F_c is the amplitude calculated from an apo structure mode in which the atoms in the figure have been omitted without any refinement. The protein N388 structure (thick bonds) is superimposed on the KF structure (thin bonds). (b, middle) The 3′-terminal two nucleotides, two metal ions, metal ligands, and Lys-299 and Glu-288 shown in stereo. (c, bottom) The remaining interactions between the protein and the trinucleotide that are not shown in Figure 5b.

additional carboxylate (Glu-288) in the vicinity of the scissile phosphate in the T4 enzyme. Which, if any, of these rather modest structural differences between these two enzymes contribute to the approximately 10^3-fold higher exonuclease activity of the T4 enzyme is not clear from the present studies.

ACKNOWLEDGMENT

We thank Jim Karam and A. K. M. Sattar for sharing unpublished results with us. The coordinates and structure factor amplitudes have been deposited for immediate release in the Brookhaven Protein Data Bank with entry numbers 1NOY, 1NOZ, R1NOYSF, and R1NOZSF.

REFERENCES

Beese, L. S., & Steitz, T. A. (1991) *EMBO J. 10*, 25−33.
Bernad, A., Blanco, L., Lázaro, J. M., Martin, G., & Salas, M. (1989) *Cell 59*, 219−228.
Blanco, L., Bernad, A., & Salas, M. (1992) *Gene 112*, 139−144.
Bloom, L. B., Otto, M. R., Eritja, R., Reha-Krantz, L. J., Goodman, M. F., & Beechem, J. M. (1994) *Biochemistry 33*, 7576−7586.
Brünger, A. T. (1992) *X-PLOR version 3.1. A System for X-ray crystallography and NMR*, Yale University Press, New Haven, CT.
Capson, T. L., Peliska, J. A., Kaboord, B. F., Frey, M. W., Lively, C., Dahlberg, M., & Benkovic, S. J. (1992) *Biochemistry 31*, 10984−10994.
Cowart, M., Gibson, K. J., Allen, O. J., & Benkovic, S. J. (1989) *Biochemistry 28*, 1975−1983.
Davies, J. F., II, Almassy, R. J., Hostomska, Z., Ferre, R. A., & Hostomsky, Z. (1994) *Cell 76*, 112−1133.
Esteban, J. A., Soengas, M. S., Salas, J., & Blanco, L. (1994) *J. Biol. Chem. 269*, 31946−31954.
Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R., & Steitz, T. A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8924−8928.
Frey, M. W., Nossal, N. G., Capson, T. L., & Benkovic, S. J. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 2579−2583.
Gibbs, J. S., Weishart, K., Digard, P., deBruynkops, A., Knipe, D. M., & Coen, D. M. (1991) *Mol. Cell. Biol. 11*, 4786−4795.
Hsu, T. (1992) Ph.D. Thesis, Tulane University, New Orleans, LA.
Huang, W. H., & Lehman, I. R. (1992) *J. Biol. Chem. 247*, 3139−3146.
Hughes, M. B., Yee, A. M., Dawson, M., & Karawa, J. (1987) *Genetics 115*, 393−403.
Jones, N. H., Cowan, S., Zhou, J. Y., & Keldgaard, M. (1991) *Acta Crystallogr. A46*, 110−119.
Joyce, C. M., & Steitz, T. A. (1994) *Annu. Rev. Biochem. 63*, 777−822.
Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W., & Steitz, T. A. (1995) *Nature 376*, 612−616.
Jancarik, J., & Kim, S. H. (1991) *J. Appl. Crystallogr. 24*, 409−411.
Kleywegt, G. J., & Jones, T. A. (1994) *Collaborative Computing Project, Number 4 (ccp4)*, 59−66.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 140−149.
Lin, T. C., Karam, G., & Konigsberg, W. H. (1994) *J. Biol. Chem. 269*, 19286−19294.
Matthews, B. W. (1968) *J. Mol. Biol. 33*, 513.
Morrison, A., Bell, J. B., Kunkel, T. A., & Sugino, A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9473−9477.
Nonay, R. L., & Reha-Krantz, L. J. (1993) *J. Biol. Chem. 268*, 27100−27108.
Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., & Steitz, T. A. (1985) *Nature 313*, 762−766.
Otwinowski, Z. (1991) ML-pHARE ccp4 Proceedings, pp 80−88, Daresbury Laboratory, U.K.
Pavlov, A. R., & Karam, J. D. (1994) *J. Biol. Chem. 269*, 12968−12972.
Pisani, F. M., & Rossi, M. (1994) *J. Biol. Chem. 269*, 7887−7892.
Reha-Krantz, L. J., Stocki, S., Nonay, R. L., Dimayuga, E., Goodrich, L. D., Konigsberg, W. H., & Spicer, E. K. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2417−2421.
Sattar, A. K. M., Lin, T. C., & Konigsberg, W. H. (1996) *Biochemistry* (submitted for publication).
Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H., & Kraut, T. (1994) *Science 246*, 1930−1935.
SERC (1979) *CCP4*, Daresbury Laboratory, Warrington, U.K.
Soengas, M. S., Esteban, J. A., Lazaro, J. M., Bernad, A., Blasco, M. A., Salas, M., & Blanco, L. (1992) *EMBO J. 11*, 4227−4237.
Spicer, E. K., Rush, J., Fung, C., Reha-Krantz, L. J., Karam, J. D., & Konisgberg, W. H. (1988) *J. Biol. Chem. 263*, 7478−7486.
Wang, T. S.-F., Wong, S. W., & Korn, D. (1989) *FASEB J. 3*, 14−21.
Xuong, N. H., Sullivan, D., Nielsen, C., Hamlin, R. C., & Anderson, E. H. (1985a) *Acta Crystallogr. B41*, 267−269.
Xuong, N. H., Sullivan, D., Nielsen, C., Hamlin, R. C., & Anderson, E. H. (1985b) *J. Appl. Crystallogr. 18*, 342−350.
Zhang, K. Y. J. (1993) *Acta Crystallogr. D49*, 213−222.

BI960178R