The EMBO Journal Vol.17 No.14 pp.4101–4113, 1998

Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme

David Jeruzalmi[1,2] and Thomas A.Steitz[1,3,4]

The Howard Hughes Medical Institute, [1]Department of Molecular Biophysics and Biochemistry and [3]Department of Chemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520-8114, USA

[2]Present address: Box 3, Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

[4]Corresponding author

The T7 RNA polymerase–T7 lysozyme complex regulates phage gene expression during infection of *Escherichia coli*. The 2.8 Å crystal structure of the complex reveals that lysozyme binds at a site remote from the polymerase active site, suggesting an indirect mechanism of inhibition. Comparison of the T7 RNA polymerase structure with that of the homologous pol I family of DNA polymerases reveals identities in the catalytic site but also differences specific to RNA polymerase function. The structure of T7 RNA polymerase presented here differs significantly from a previously published structure. Sequence similarities between phage RNA polymerases and those from mitochondria and chloroplasts, when interpreted in the context of our revised model of T7 RNA polymerase, suggest a conserved fold.

Keywords: crystal form/polymerase domain/T7 RNA polymerase/T7 lysozyme/transcriptional inhibition

Introduction

While there is now extensive structural information on many DNA polymerases and their substrate complexes (Arnold *et al*., 1995; Brautigam and Steitz, 1998), much less is known of the structural basis of DNA-dependent RNA polymerase function. Although the chemistry of polymerization is likely to be conserved (Delarue *et al*., 1990; Steitz, 1993), RNA polymerases perform several important functions not found in DNA polymerases (von-Hippel *et al*., 1984; Erie *et al*., 1992). First, they are able to initiate RNA synthesis without requiring a primer oligonucleotide; moreover, these enzymes must recognize a specific duplex promoter DNA sequence for efficient initiation. Secondly, RNA polymerases exhibit an abortive initiation phase, characterized by the synthesis of short RNA products, during which the position of the promoter on the polymerase remains fixed (Ikeda and Richardson, 1986). Only after undergoing a transition of unknown nature does the polymerase enter the processive elongation phase of RNA synthesis. Thirdly, RNA polymerases are the targets of a host of regulatory proteins (activators, inhibitors, terminators and anti-terminators) that modulate gene expression (Tjian, 1996).

The RNA polymerase (RNAP) from bacteriophage T7 has several advantages for use in the study of the structural basis of the unique properties of RNA polymerases. At 98 kDa, it is smaller than the multi-subunit prokaryotic and eukaryotic RNA polymerases, facilitating X-ray crystallographic studies. Its well-characterized interactions with promoter DNA (Ujvari and Martin, 1997) and the inhibitor T7 lysozyme (Kumar *et al*., 1997; Zhang and Studier, 1997) enable a detailed structure–function analysis. Homology between T7 RNA polymerase and the well-studied DNA polymerase I family (Brautigam and Steitz, 1998; Doublie *et al*., 1998) is important for deciphering the structural basis of the functional differences between RNA and DNA polymerases.

The bacteriophage T7 RNA polymerase–T7 lysozyme (PL) complex participates in the regulation of gene expression upon infection of *Escherichia coli*. During infection, the phage RNA polymerase directs transcription of two classes of viral genes (II and III). As levels of the viral class II gene-product T7 lysozyme (LYS) rise, RNAP molecules are sequestered into transcription-inhibited PL complexes (McAllister and Wu, 1978; Moffat and Studier, 1987). Since phage T7 promoter DNA sequences display varying efficiency at RNA chain initiation (class III $\gg$ class II), the resulting reduction in available phage transcriptional capacity shuts off expression of class II genes (McAllister *et al*., 1978, 1981). Transcription of class III gene products continues until lysis (Studier, 1972). Control of phage transcription in this way achieves a physiological economy of gene products relevant to the phage life-cycle (Studier and Dunn, 1983). Consistent with this scheme, phage lacking the lysozyme gene or with an RNAP resistant to lysozyme inhibition, fail to repress class II gene expression (McAllister *et al*., 1981; Moffat and Studier, 1987).

While the PL complex is competent to catalyze the synthesis of short RNA molecules, it fails to clear the abortive initiation phase. Early in its transcriptional cycle, RNAP forms an unstable abortively initiated complex with promoter DNA and synthesizes short (8–10 base) RNA products (Martin *et al*., 1988; Ling *et al*., 1989). Transition to a stable, processive elongation complex occurs by an uncharacterized reorganization of the ternary complex (RNAP–DNA–RNA) that may include protein conformational changes as well as the establishment of interactions between nascent RNA and the N-terminal domain (residues 1–325) of RNAP (Muller *et al*., 1988). Recent work has demonstrated that T7 lysozyme, a phage-encoded protein (M_r ~17 kDa) with both a peptidase (Inouye *et al*., 1973) and transcriptional inhibition activity (Moffat and Studier, 1987; Ikeda and Bailey, 1992), prevents this transition (Kumar and Patel, 1997; Zhang and Studier, 1997). Similar to the unstable abortive complex, the PL complex can only synthesize short RNA

products (up to 15 bases) despite making near wild-type interactions with promoter DNA, RNA and nucleotides. Interestingly, once RNAP has entered the elongation phase, it is no longer sensitive to lysozyme inhibition (Zhang and Studier, 1997).

The structure of the T7 RNA polymerase–T7 lysozyme complex described here reveals that T7 lysozyme binds to RNAP below the fingers sub-domain at a site remote from the polymerase active site, establishing that inhibition is not achieved by steric blocking of the active site. The surface of LYS that contacts RNAP accurately reflects genetic and biochemical analyses (Cheng *et al*., 1994). Our model for RNAP agrees with that of Sousa *et al*. (1993) (PDB entry: 2RNP) only with respect to overall shape and location of secondary-structure elements, but differs in the positioning of the amino acid sequence onto tertiary structure.

Results

Structure determination

We have determined the structure of the PL complex in two crystal forms (designated I and III) containing three and one complexes in the asymmetric unit, respectively, and giving four independent views. Electron density maps, calculated using multiple isomorphous replacement phases in each crystal form (Table I), were used to position the complex(es) by molecular replacement. Our model (Figure 1) was built into solvent-flattened and cross-crystal domain averaged maps (Kleywegt and Read, 1998) and refined using the CNS program (Brunger, 1998) (Materials and methods, and Figure 2). Discussion of the structure will focus exclusively on the PL complex in crystal form III which has been refined to an R_{free} = 31.8% (30–2.8 Å).

Overall architecture

The structure of the PL complex reveals an irregularly-shaped protein assembly with lysozyme binding to a site distant from the polymerase active site (Figure 1). RNAP consists of a polymerase (residues 326–883) and an N-terminal domain (residues 1–325), whose boundary we have re-defined based on this structure and on its comparison with other pol I polymerase domains (Ollis *et al*., 1985). The polymerase domain can be divided into sub-domains denoted thumb, fingers and palm, named as a structural analogy to the right hand. These sub-domains have been observed in all nucleic acid polymerase families with structures in the protein database (Ollis *et al*., 1985; Kohlstaedt *et al*., 1992; Pelletier *et al*., 1994; Wang *et al*., 1997). The polymerase active site, identified by conserved residues whose mutation disrupts catalytic activity (Bonner *et al*., 1992; Osumi-Davis *et al*., 1992), resides within the deep cleft formed by these three sub-domains. The N-terminal domain, whose disruption severely reduces processive RNA synthesis (Muller *et al*., 1988; He *et al*., 1997), is located in front of the palm and thumb sub-domains.

The polymerase–lysozyme complex

T7 lysozyme binds to RNAP on the surface opposite to the active-site cleft (Figures 1 and 3), suggesting an indirect mechanism of inhibition. Lysozyme binding therefore, leaves the active-site cleft open to interaction with DNA, nucleoside triphosphates, and single-stranded RNA, as predicted after biochemical studies (Zhang and Studier, 1997). The interface is composed of polar and hydrophobic contacts, and buries ~2100 Å^2 of solvent-accessible surface area. The RNAP portion of the lysozyme binding site includes structural elements from the palm (the extended foot module) and fingers sub-domains, and the N-terminal domain (Figure 3A and B). Binding of lysozyme effectively fixes the positions of these elements relative to each other, possibly altering their orientation in comparison with that of free polymerase or restricting conformational changes that may be required during various stages of the transcriptional cycle (Mookhtiar *et al*., 1991; Sastry and Hearst, 1991; Sousa *et al*., 1992). Either or both of these may be relevant to the mechanism of inhibition. Although complicated by errors in connectivity and amino acid register in the previous structure (Table II), comparison of our model with that of uncomplexed RNAP (Sousa *et al*., 1993) shows that the position of the fingers sub-domain in the PL complex is translated by ~4–5 Å towards the palm sub-domain (Figure 3D).

The locations of RNAP mutations that result in resistance or hyper-sensitivity to lysozyme inhibition fall into two distinct classes (Figure 3C) that provide differing insights into possible mechanisms of inhibition. Mutations in RNAP, selected by propagating T7 phage in a background of higher than wild-type concentration of T7 lysozyme, confer resistance to inhibition and map to the lysozyme binding site. In this case, resistance is probably a consequence of a weakened polymerase–lysozyme interaction. By contrast, mutations selected against a LYS mutant of reduced inhibitory potency result in RNAP hypersensitivity to lysozyme inhibition (Moffat and Studier, 1987; Zhang, 1995). These mutations lie away from the protein–protein interface and cluster in the fingers and palm sub-domains. Perhaps changes in these residues, the majority of which are not surface accessible, alter the structure or dynamics of RNAP in the affected regions, enhancing its ability to respond to binding of lysozyme.

LYS, although only a small protein, fulfills two diverse activities using different portions of the molecule. Binding to RNAP does not significantly alter the structure of lysozyme (Cheng *et al*., 1994) (PDB entry:1LBA) and its contact surface agrees with the results of genetic and biochemical studies (Figure 3B). Residues 2–6 are disordered in the uncomplexed structure but are visible in our electron density maps and form the polymerase binding surface along with helix αA (residues 29–39). Access to the peptidase active center is occluded due to binding of LYS to RNAP, explaining mutual exclusivity of the two activities (Cheng *et al*., 1994). Indeed, preparation of crystal form III requires displacement of the active-site zinc ion by mercury, thereby disrupting the active site by coordinating to a different set of ligands (C18 and C130). This suggests that a functional peptidase activity is superfluous for binding to RNAP.

Structural segregation of the two activities of lysozyme

Table I. Crystallographic structure determination and refinement

A. Multiple isomorphous replacement

Crystal form III; space group C2: a = 270 Å, b = 93 Å, c = 63 Å, α = 90°, β = 103°, γ = 90° (T = –175°C)

Data set/type	Resolution	Reflections	R_{sym} %	R_{iso} %	Phasing power
Native	20–3.3 Å	22 047 97.2%	6.0	–	–
N7-platinated guanosine	20–3.5 Å	18 924 96.1%	8.0	14.1	0.81
PIP1	20–5.0 Å	5641 78.7%	9.3	30.3	1.29
Na•Au(CN)$_2$	20–3.0 Å	27 537 86.9%	10.6	12.9	0.75
para-hydroxy-mercuri-benzoate (form IIIa)	20–2.8 Å	38 337 93.4%	4.2	34.6	N/A
bis (acetoxy-mercuri) toluidine	20–4.0 Å	11 600 93.2%	8.3	22.7	1.12
PIP2	20–4.0 Å	12 612 94.1%	10.3	27.8	0.82
5′-SH mercurated guanosine	20–5.0 Å	18 812 94.2%	10.0	17.9	0.90

Overall figure of merit: 0.37

Crystal form I; space group C2: a = 320 Å, b = 93 Å, c = 229 Å, α = 90°, β = 129°, γ = 90° (T = –175°C)

Data set/type	Resolution	Reflections	R_{sym} %	R_{iso} %	Phasing power
Native	20–3.3 Å	69 094 87.2%	8.2	–	–
PIP	20–4.0 Å	25 801 52.0%	9.7	41.0	0.92
5′-SH mercurated guanosine	20–4.0 Å	36 671 60.0%	8.4	19.8	0.82
Uranyl acetate	20–3.7 Å	45 448 80.9%	9.0	14.0	0.92
Na•Au(CN)$_2$	20–3.4 Å	53 009 80.1%	7.1	21.1	0.88

Overall figure of merit: 0.28

B. Density modification

R-factor (I):	26.5%	Correlation coefficient:	0.88
R-factor (III):	20.1%	Correlation coefficient:	0.93

C. Refinement

Data set	Resolution	Reflections F> 2σ	Total atoms (No.)	$R_{working}$ (R_{free})	r.m.s. D bonds	r.m.s. D angles	r.m.s. B for values
FormIIIa pHMBS soak	30.0 – 2.8 Å	32 864	7726	26.33 (31.83)	0.012 Å	1.80°	1.89 Å^2

$R_{sym} = \Sigma|(I - <I>)|/\Sigma I$, where $<I>$ is integrated intensity averaged over symmetry equivalents.
$R_{iso} = \Sigma|(F_{PH} - F_P)|/\Sigma F_P$, where F_P and F_{PH} are the native and derivative structure factor amplitudes, respectively.
Phasing power = $[\Sigma|F_{PH(calc)}|^2/\Sigma |F_{PH(obs)} - F_{P(calc)}|^2]^{1/2}$.
Figure of merit = cosine $<\sigma(\Delta\phi)>$.
PIP: di-μ-iodobis-(ethylenediamine) diplatinum (II) nitrate.
R-factor = $\Sigma|[F_{(obs)} - F_{(calc)}]|/\Sigma F_{(obs)}$.

$$\text{Correlation coefficient} = \frac{[\Sigma(F_{(obs)} - <F_{(obs)}>)\,(F_{(calc)} - <F_{(calc)}>)]}{[\Sigma(F_{(obs)} - <F_{(obs)}>)^2\Sigma(F_{(calc)} - <F_{(calc)}>)^2]^{1/2}}$$

$F_{(calc)}$ represents structure factors obtained from back-transformation of solvent-flattened/NCS averaged maps.
$R_{value} = \Sigma|(F_P - F_{P(calc)})|/\Sigma F_P$. R_{free} is the R_{value} for 10% of the reflections that were omitted from the refinement.

is also indicated by sequence similarity (Figure 5B) between T7 lysozyme and the lysozyme from *Haemophilus influenzae* (HINF) (Fleischmann *et al*., 1995). The majority of residues implicated in peptidase activity reside within the C-terminal 100 residues of T7 lysozyme (Cheng *et al*., 1994) that are held in common between

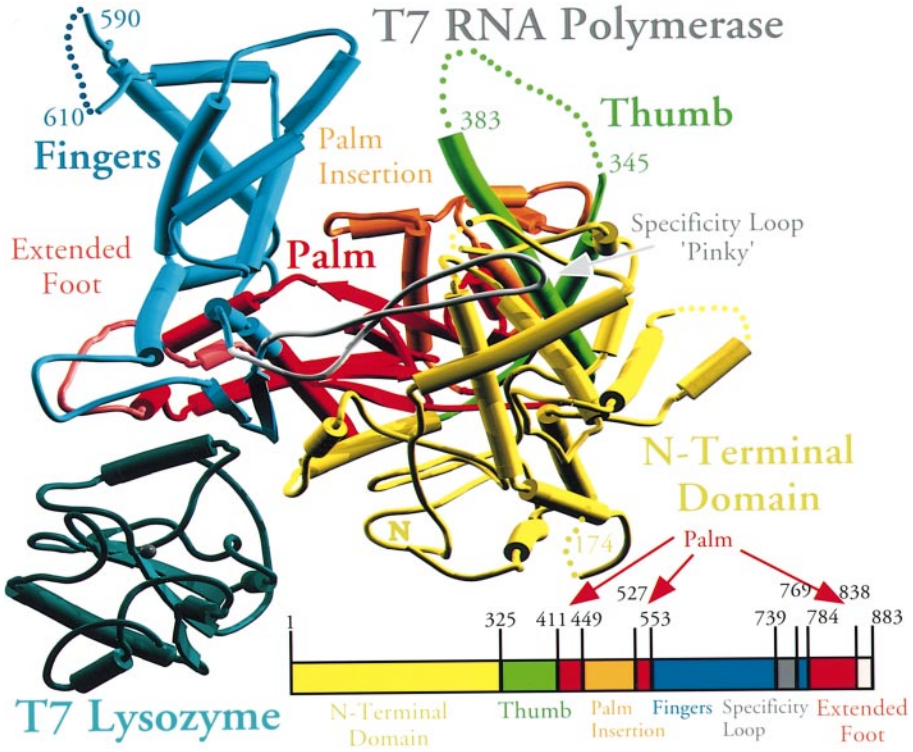

Fig. 1. Structure of the T7 RNA polymerase–T7 lysozyme complex. Schematic representation of the PL complex, in which α-helices are depicted as tubes and β-strands as arrows. This representation is colored by domain, sub-domain or module, with the N-terminal domain (8–325), yellow, the thumb (326–411), green, the palm (412–449, 528–553, 785–879), red, the palm insertion module (450–527), orange, the fingers (554–739, 769–784), blue, the 'pinky' specificity loop (740–769), white, extended foot module (838–879), pink and T7 lysozyme (blue-green) with bound mercury atom in silver. Disordered portions of RNAP are indicated by colored dots. The location of the various domains, sub-domains, or modules, colored as above, is projected onto the primary sequence of T7 RNA polymerase, represented as a bar.

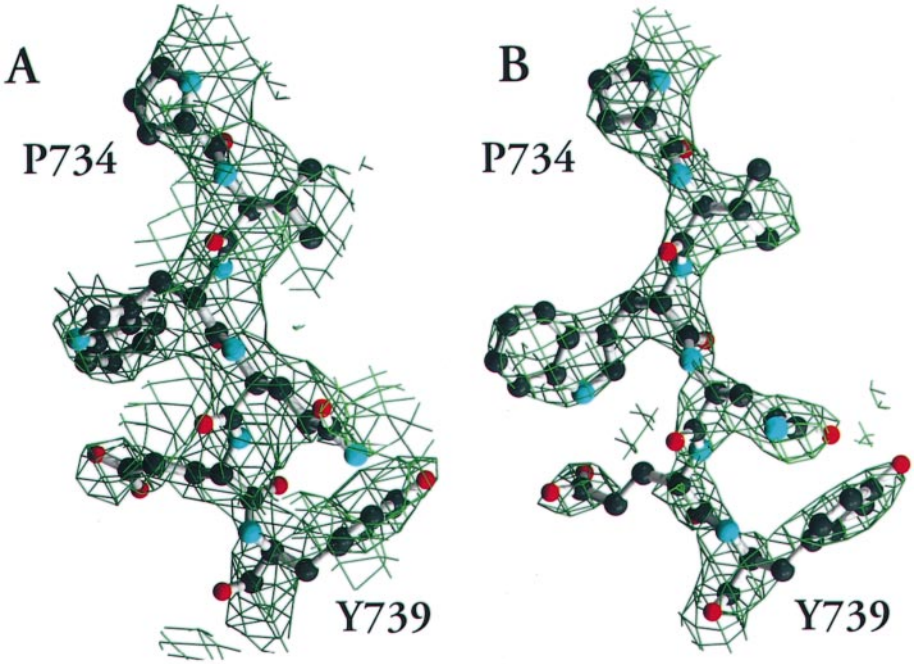

Fig. 2. Representative electron density maps. (**A**) Residues 734–739 of an intermediate model of the PL complex are superimposed onto the final 4-fold domain averaged electron density map (20–3.0 Å) in crystal form III. The map is contoured at 1.3σ. (**B**) An annealed $2F_o–F_c$ omit electron density map, calculated by excluding residues 734–739 of the final PL complex model. The map is contoured at 1.3σ.

the two enzymes (Figure 5B). An insertion near the C-terminal of the HINF lysozyme, not in the phage protein, could provide the residues missing in the HINF enzyme (e.g. LYS H17, Y46), which include ligands for the catalytically important zinc ion. The N-terminal 52 residues of LYS, missing in the HINF enzyme, contain residues important for RNAP interaction and transcriptional inhibition (Figure 3B).

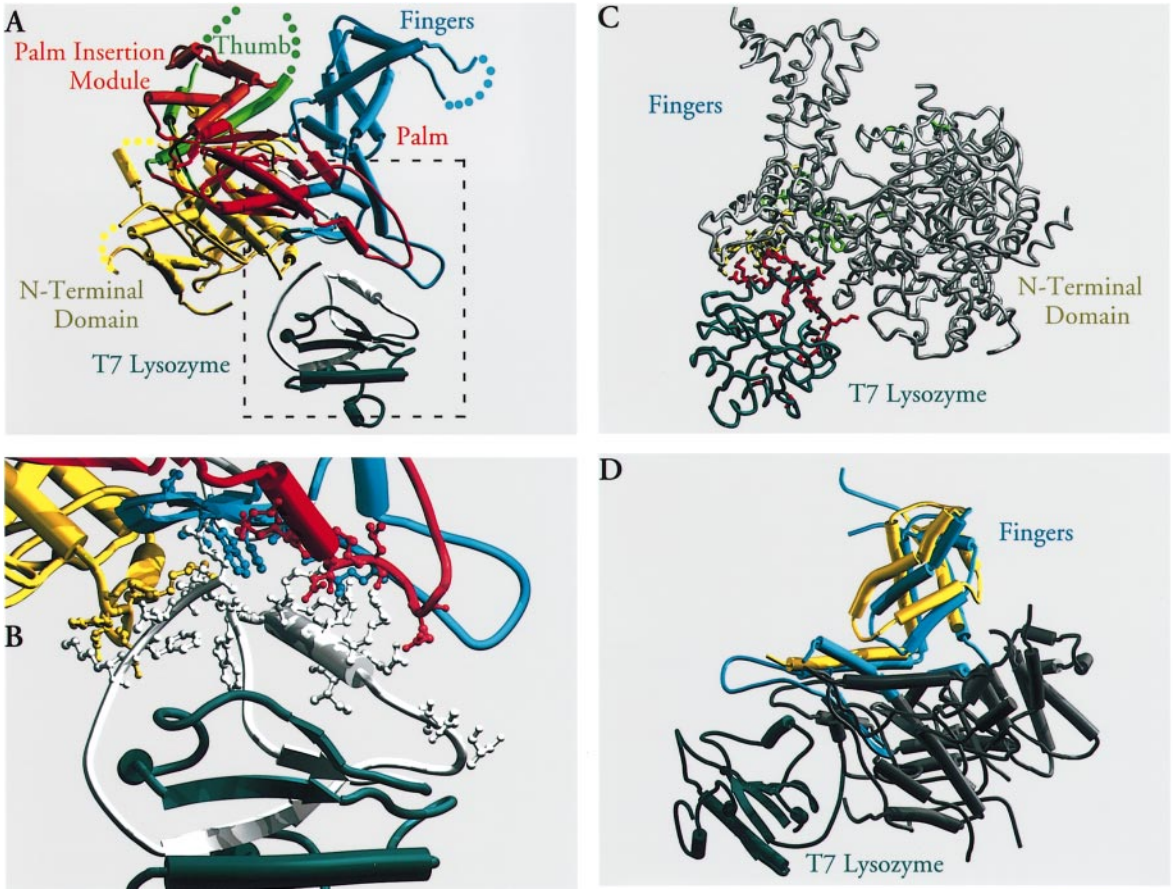

Fig. 3. Interaction between T7 RNA polymerase and T7 lysozyme. (**A**) Detailed view of the T7 lysozyme binding site on RNAP reveals that the interaction surface is composed of elements from the N-terminal domain and the palm and fingers sub-domains, each colored as in Figure 1. T7 lysozyme is colored blue-green, except for its RNAP interaction domain (amino acids 2–52) which is in white. This orientation of the complex is orthogonal to that of Figure 1 about the vertical axis. (**B**) A close-up view of the T7 lysozyme binding site in which residues within 4.0 Å of the interface are depicted by 'ball and stick' representation. RNAP residues in contact with LYS are colored (as in Figure 1) by the domain in which they reside. (**C**) Mapping of inhibition-affecting mutations in RNAP and LYS onto the structure of the PL complex. The complex is represented as a set of connected C_α atoms. RNAP is colored grey and LYS is blue-green. Mutations in RNAP that affect interaction with T7 lysozyme segregate into two classes: mutants that confer resistance to lysozyme (colored yellow); and those that confer hypersensitivity to lysozyme inhibition (colored green) (Moffat and Studier, 1987; Zhang, 1995). Residues in T7 lysozyme, identified by genetic studies affecting interaction with RNAP, are colored red (Cheng *et al*., 1994). Orientation of the complex is as in Figure 1. (**D**) Super-position of RNAP from the PL complex onto the the unliganded RNAP structure (amino acids 530–539, 806–817). In comparison with uncomplexed RNAP, the fingers sub-domain of the PL complex is translated 4–5 Å towards the palm sub-domain. The fingers sub-domain of the RNAP from the PL complex is drawn in light blue, with helices as tubes and strands as arrows. The rest of the RNAP from the PL complex is colored grey and for clarity, portions of its N-terminal domain and thumb sub-domain have been deleted. Only the fingers sub-domain of the unliganded RNAP is shown, colored yellow. The orientation of the complex is rotated 90° about the vertical axis from that of Figure 1.

Table II. Differences between the revised RNAP structure and previous structure (PDB code 2RNP; Sousa *et al*., 1993)

Type of difference	Location in the model
Chain direction	8–164
Connectivity	177–294
Residue frameshifts:	
1–3 positions	293–345
1 position	505–533
1–2 positions	538–562
4–7 positions	721–765
1–3 positions	838–860

The RNA polymerase domain

The structure of RNAP in the PL complex is similar to the model published by Sousa *et al*. (1993) with respect to overall shape and location of secondary-structure elements. However, our model diverges significantly with respect to the mapping of amino acid sequence to tertiary structure (summarized in Table II). The polymerase domain has a U-shape fold, characteristic of all pol I polymerases (Steitz, 1993; Arnold *et al*., 1995; Brautigam and Steitz, 1998). It contains not only the thumb, palm and fingers sub-domains (Figure 1), but also segments unique to phage-type RNA polymerase that fulfill specialized functions (Muller *et al*., 1988; Raskin *et al*., 1992; Gardner *et al*., 1997).

The thumb sub-domain

Amino acids 326–411 comprise the thumb sub-domain, which forms a wall around the right side of the catalytic cleft (Figure 1). The top portion (345–383) of the thumb sub-domain is disordered in the PL structure. As a result of crystal contacts in free RNAP, it is a long helix that rises above the palm sub-domain (Sousa *et al*., 1993). Poorly ordered thumb sub-domains are a common feature in polymerases in the absence of nucleic acid substrates or stabilizing crystal contacts (Ollis *et al*., 1985; Jacobo-

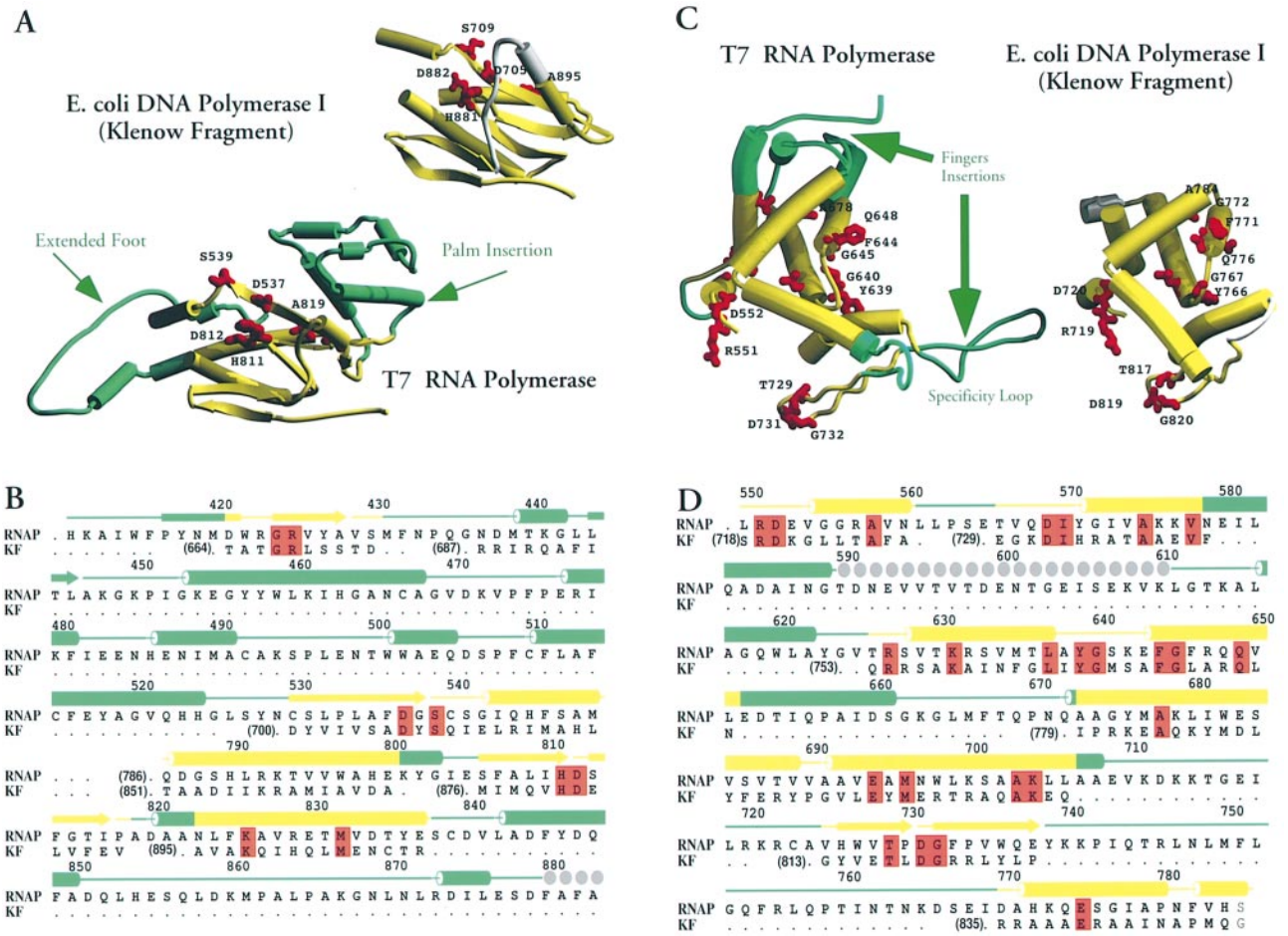

Fig. 4. Alignment of T7-RNA polymerase onto the Klenow fragment of *E.coli* DNA polymerase I. (**A**) A structural comparison of the palm sub-domains of RNAP and KF, identically oriented after super-position of homologous secondary structure elements (colored yellow). Polypeptide segments unique to phage-type RNA polymerases are colored light green. The residue side chains, identical in both structures based upon the structural alignment, are shown in red. (**B**) Alignment of primary sequences of the palm sub-domains of RNAP and the KF, based upon the structural super-position. Residues within common secondary-structure elements of the two structures are colored in yellow; identical residues between the two superimposed structures are shown in red. The secondary structure of the revised model of RNAP is colored as in (A). (**C**) A structural comparison of the fingers sub-domains of RNAP and the KF fragment. Again, regions that superimpose are colored yellow and identical residues are shown in red. (**D**) Primary sequence alignment of the fingers sub-domains of RNAP and the KF, based upon the structural super-position in (C).

Molina *et al*., 1993; Sousa *et al*., 1994; Kim *et al*., 1995). These domains are observed to be in contact with the backbone, spanning the minor groove of primer–template duplexes (Eom *et al*., 1996; Doublie *et al*., 1998; Kiefer *et al*., 1998). Consistent with this role, mutant RNAPs with shorter thumbs are less processive, presumably due to a lower affinity for template DNA (Bonner *et al*., 1994b). This is distinct from disruptions to the N-terminal domain that cause loss of processivity due to reduced affinity for RNA (Muller *et al*., 1988; He *et al*., 1997). Analogous constructs in the Klenow fragment (KF) of *E.coli* DNA polymerase I, containing abbreviated thumb sub-domains, display similar defects (Minnick *et al*., 1996).

The palm sub-domain

The palm sub-domain is located at the base of a deep cleft, bounded by the fingers and thumb sub-domains, and contains residues 412–553 and 785–879 (Figure 1). The C-terminal four residue stretch of RNAP (880–883), which is known to be flexible in solution (Mookhtiar *et al*., 1991), is poorly ordered in our electron density maps. Located within the palm sub-domain is a trio of β-strands whose structure is conserved in every nucleic acid polymerase except DNA polymerase β (Steitz *et al*., 1994; Brautigam and Steitz, 1998). Equivalent C_α atoms between RNAP and the KF within this segment, which include sequence motifs A and C (Delarue *et al*., 1990), can be superimposed with an r.m.s. fit of 2.1 Å (Figure 4A and B). This constellation of β-strands harbors a pair of absolutely conserved aspartate residues (D537, D812; Figure 5A) proposed to orient two metal ions for catalysis of the polymerase reaction (Steitz, 1993; Joyce and Steitz, 1994). This proposal has recently received experimental confirmation in the case of the pol I family of DNA polymerases (Doublie *et al*., 1998). Consistent with this proposed role, mutation of RNAP residues D537 and D812 severely disrupts catalysis, without affecting its affinity for rNTPs (Bonner *et al*., 1992; Osumi-Davis *et al*., 1992). Besides playing a role in catalysis, palm sub-domains have recently been implicated in distinguishing between ribo- and 2′-deoxynucleotides (Gao *et al*., 1997; Joyce, 1997). MuLV reverse transcriptase uses a bulky hydrophobic residue (F115) to sterically block the 2′-OH

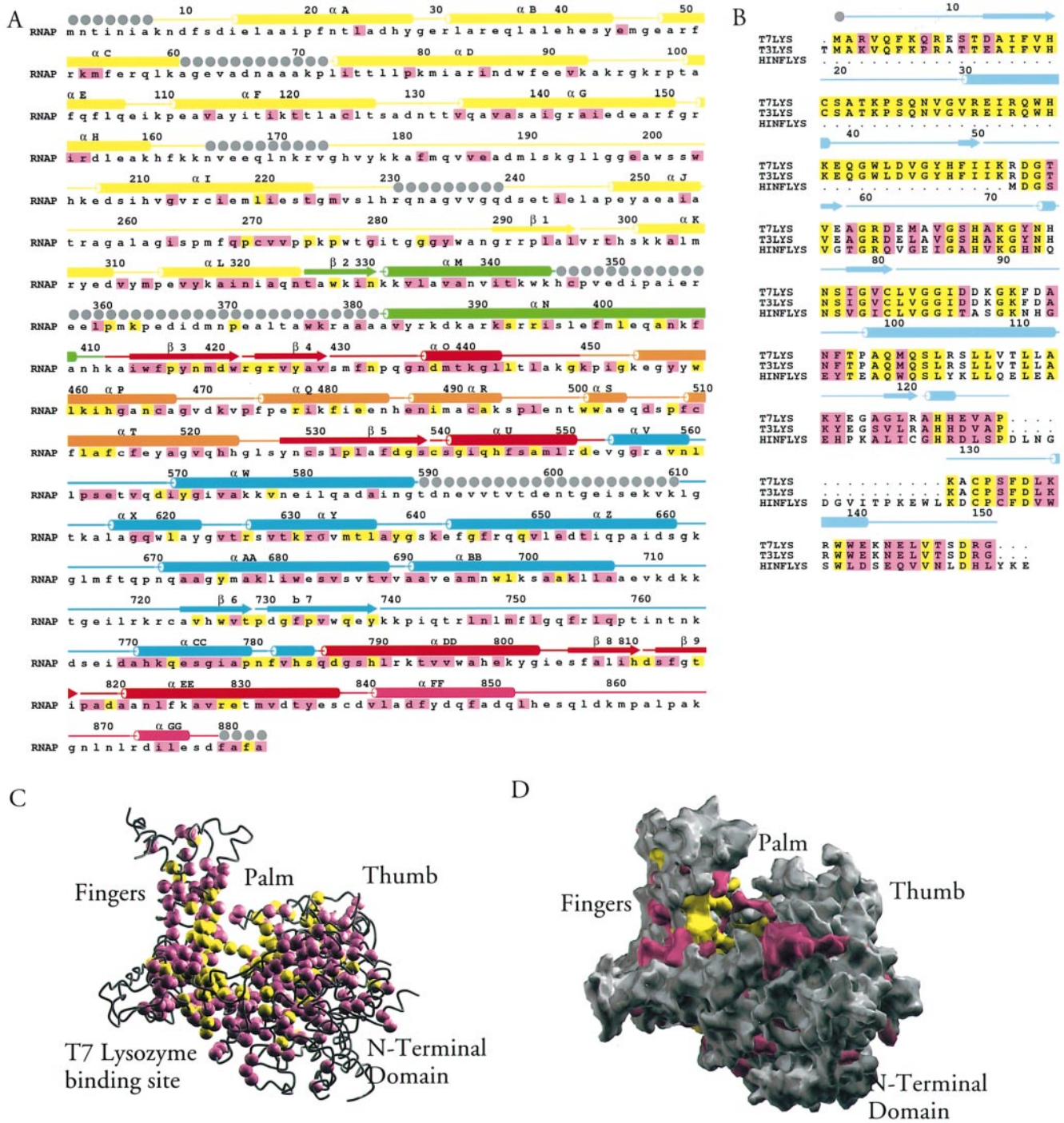

Fig. 5. Sequence alignments of phage-type RNA polymerases and lysozymes. (**A**) Multiple sequence alignment of phage-type RNA polymerases are projected onto the primary sequence of T7 RNA polymerase. Identicalities are colored yellow and conservative substitutions as defined in Livingstone and Barton (1993) are pink. RNAP sequences from phage T7, mitochondria of *Saccharomyces cerevisiae*, human mitochondria and from the chloroplast of *Arabidopsis thaliana* are included in this alignment. The alignment is similar to that of a ClustalW alignment of primary sequences from 23 phage-type RNA polymerases from phage T7 (accession No. 431189), T3 (133452), Sp6 (133451), K11 (133450), complete or partial human mitochondrial sequences (2114396), yeast (172490), *Arabidopsis thaliana* (2326363), *Cenopodium album* (1865721), *Oryza sativa* (1236334), *Pynococcus* (1236336), *Tetrahymena pyr.* (1236348), *Thrausto chytriumaureum* (1236338), *Triticum aes*. (1236346), *Acanthamoeba castellani* (1236340), *Cryptomonas* (1236342), *Isochrysis sp*. (1236332), *Naegleria fowl* (1236344), and mitochondrial sequences of fungal senescence-inducing plasmids from *Neurospora* (283360). The secondary structure from the revised model of T7 RNA polymerase, colored by domain as in Figure 1, accompanies the alignment. Grey spheres represent portions of the molecule not visible in our electron density maps. (**B**) Multiple sequence alignment of three phage-type lysozymes from phage T7 (T7LYS), T3 (T3LYS) and *Haemophilus influenzae* (HINFLYS). The color scheme is as in (A). Accompanying the sequence alignment is the secondary structure for the model for T7 lysozyme. (**C**) Location in the structure of residues conserved among phage-type RNA polymerases. T7 RNA polymerase is depicted as a set of connected C_α atoms, colored grey. Spheres colored as in (A) depict the position of conserved residues. (**D**) Clustering of surface residues, conserved in phage-type RNAPs, to the active-site cleft. The alignment of (A) is projected onto the RNAP molecular surface as calculated in GRASP. The color scheme is the same as in (A).

of ribonucleotides, while allowing binding of deoxynucleotides (Gao *et al*., 1997). By contrast, phage-type RNA polymerases have an invariant glycine (RNAP:G542, Figure 5A) at this position which would favor binding of rNTPs.

The fingers sub-domain

The fingers sub-domain rises above the palm sub-domain forming a wall on the left of the active-site cleft and includes residues 554–784; residues 711–720 are partially ordered in our maps and have been modeled as polyalanine, while residues 590–610 are not visible. Structural comparison (Figure 4C and D) of the fingers sub-domains of RNAP and the KF reveals many regions of conserved sequence and secondary structure as well as some striking differences. The RNAP fingers sub-domain has several inserted segments relative to KF; two of these (residues 580–625 and 652–671) have the effect of making the RNAP fingers taller than the DNA polymerase counterpart (Figures 4C and 6) giving rise to a deeper active-site cleft (~30 Å versus ~20 Å). A third insertion, which we have referred to as the 'pinky' specificity loop, extends from residue 740 to 769 and is involved in promoter recognition (Raskin *et al*., 1992). The fingers sub-domains occupy different relative orientations in each of the three PL complexes in crystal form I (data not shown). The relationship between this conformational heterogeneity and lysozyme inhibition is unclear at present, but may reflect inherent RNAP flexibility. Interpretation of the results of mutational data for residues in the fingers sub-domain, in order to assign functions to particular surfaces, is difficult due to the complicated effect of the defects on the polymerase reaction (Rechinsky *et al*., 1993; Bonner *et al*., 1994a; Kostyuk *et al*., 1995; Rechinsky *et al*., 1995). Crystal structures of DNA polymerases bound to nucleic acid duplexes and dNTP demonstrate that the fingers sub-domain interacts with both the template strand as well as nucleotide (Eom *et al*., 1996; Doublie *et al*., 1998; Kiefer *et al*., 1998).

The N-terminal domain

The N-terminal domain is located in front of the polymerase domain and forms the front wall of the catalytic cleft, contributing to its concave shape (Figures 1 and 6). Our structure encompasses residues 8–325; residues 60–73, 164–174, which are known to be flexible in solution (Ikeda and Richardson, 1987; Muller *et al*., 1988), and 231–241 are not visible in our electron density maps. The greatest divergence between our structure and that published by Sousa *et al*. (1993) is seen in this domain. We find differences in main chain connectivity (177–294), amino acid register (293–345), and the N to C direction of the main chain (8–164) (Table II). A search of the structural database fails to identify proteins with similarity to the N-terminal domain (Holm and Sander, 1994).

Biochemical and mutational analyses have established

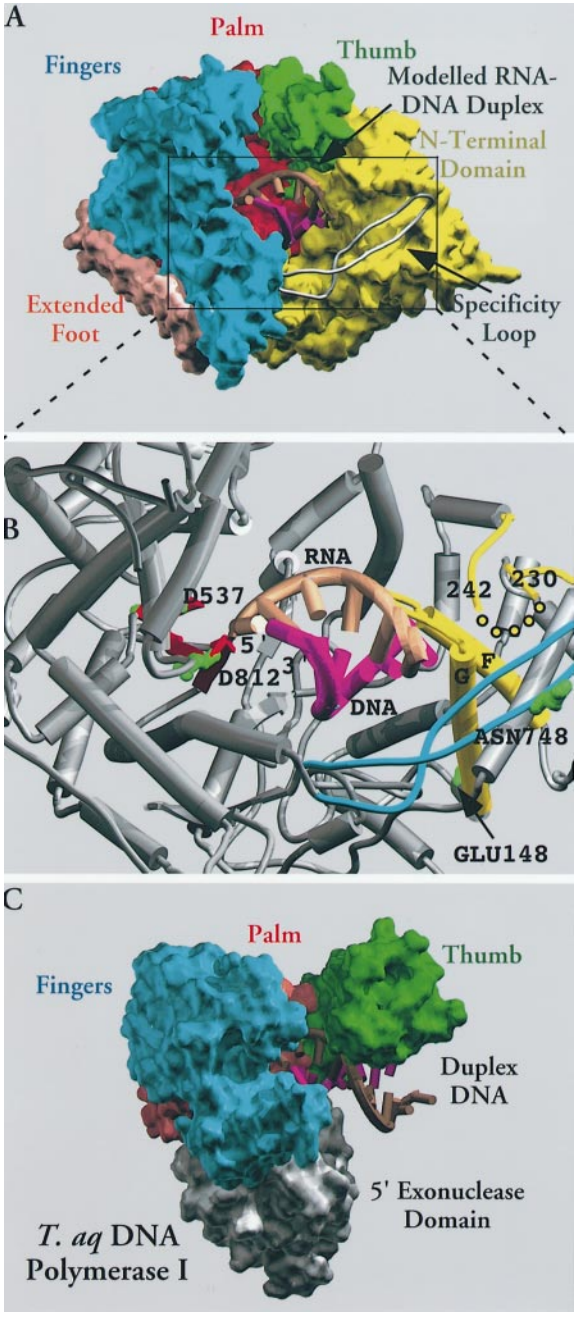

Fig. 6. Model for interaction with promoter DNA and nascent RNA. (**A**) RNAP is represented as a molecular surface, colored by domain as in Figure 1, showing the highly concave shape of its active-site cleft. The 'pinky' specificity loop is depicted as a light blue ribbon of connected C_α atoms. The nucleic acid duplex present in the RNAP cleft is represented as a ribbon (RNA, brown; DNA, magenta), corresponding to the phosphodiester backbone, with cylinders as bases. Its location derives from super-position of the three β-strands from the active site from the *Taq* DNA polymerase–duplex DNA complex (Eom *et al*., 1996). Of the 14/15mer duplex present in the DNA polymerase structure, 6–8 bases can be accomodated in the cleft of RNAP. (**B**) A close-up view of the active-site cleft found in RNAP with the modeled nucleic acid duplex showing clashes with the N-terminal domain. This view of RNAP is identical to (A) except that α-helices are depicted as tubes, β-strands as arrows. Helices F, G and the adjoining loop (N-terminal domain) are colored yellow, the active-site β-strands are in red and the specificity loop is colored light blue. For clarity, the remaining portions of the RNAP structure are colored grey. The position of residues significant for rNTP binding (GLY542), catalysis (ASP812), promoter interaction (ASN 748) are indicated. The location of GLU148, a residue whose mutation severely disrupts RNA binding and processivity (He *et al*., 1997) is highlighted. (**C**) The DNA polymerase from *Thermus aquaticus* is represented as molecular surface showing its more open active-site cleft. The surface is colored as in (A). The vestigial 3′-5′ exonuclease domain is colored in white. For clarity, the 5′ exo-nuclease domain has been deleted. The DNA from the complex is modeled and colored as in (B).

a role for the N-terminal domain in interaction with upstream regions of promoter DNA and the nascent RNA chain, leading to establishment of a processive enzyme (Ikeda and Richardson, 1987; Muller *et al*., 1988). Revision of the RNAP structure has allowed re-interpretation of the consequences of disruptions to this domain that shed new light on its role in the transcription cycle. Residues 231–241 are poorly ordered in our structure; however, we expect their location to be near the 'pinky' specificity loop (Figure 6B), a segment directly implicated in promoter DNA contacts (Raskin *et al*., 1992). The presumed location of this region and the result that mutational insertions after ASP240 abolish promoter-dependent transcription without loss of RNA polymerase activity (Patra *et al*., 1992; Sousa *et al*., 1992) strongly suggests some role in interaction with promoter DNA. A pair of solvent-accessible helices (F/G) and the adjoining loop, encompassing amino acids 125–140, form a structural element that faces towards the active-site cleft and seems well situated for interaction with promoter DNA and nascent RNA after it has cleared the transcription bubble (see section on modelling of nucleic acids and Figure 6). Site-specific mutations within the N-terminal domain with very significant defects in processivity of the polymerase reaction, binding of single-stranded RNA, and termination of elongation (He *et al*., 1997) map directly below this element. Many of these residues are not solvent accessible and their location does not appear to support a role in contacting nucleic acids. Processive elongation of RNAP is also reduced by endo-proteolysis between residues 170 and 180 of the N-terminal domain, while complete removal of this domain prevents the transition to processive elongation altogether and abolishes RNA binding *in vitro* (Muller *et al*., 1988). Although this segment is disordered in our structure, its presumed location on the opposite surface to that facing the catalytic cleft (Figures 1 and 6) seems inconsistent with the observed defects. We speculate that these two types of disruptions to the N-terminal domain as well as other site-directed mutations (Gross *et al*., 1992; Lyakhov *et al*., 1997) may compromise its structural integrity, preventing it from fulfilling its critical role in the transcription cycle.

Structural differences between pol I family RNA and DNA polymerases

Comparison of the RNAP structure with those of homologous pol I family DNA polymerases (Figure 4) reveals the presence of unique segments, some of which fulfill functions specific to RNA polymerases. The main structural differences between a pol I family DNA and RNA polymerase include a lack of any exonuclease domain in RNAP, the presence of palm and fingers sub-domain insertions, and addition of the N-terminal domain. The palm sub-domain of RNAP contains a 77 residue insertion between amino acids 450 and 527, called the palm insertion module (Figures 1 and 4A and B). This compact domain rises above and behind the active site and makes contacts with the fingers sub-domain, effectively closing off the back of the putative nucleic acid-binding channel. Its conservation (Figure 5A) in phage-type RNA polymerases implies an important function that remains obscure in the absence of published mutational data. A search of the structural database fails to identify homologous structures (Holm and Sander, 1994).

A second insertion, relative to DNA polymerases, which we have termed the extended foot module, begins at residue 838 and includes the C-terminal tetrapeptide F880–A881–F882–A883 referred to as the 'foot' (Mookhtiar *et al*., 1991). This insertion is located at the front edge of the catalytic cleft, and extends beneath the fingers sub-domain to a location behind the active site, positioning the C-terminus behind the palm insertion module (Figure 3A). The foot, whose flexibility in solution is modulated by binding of promoter DNA (Mookhtiar *et al*., 1991), is disordered in the PL complex. In the uncomplexed RNAP structure, the foot is modelled in an extended conformation, packed below one of the active-site β-strands (Sousa *et al*., 1993). The observed location of the last visible residue (879) in our structure is just behind the active site, close enough to allow positioning of the foot such that it could fulfill its proposed roles of interaction with incoming rNTPs and promoter DNA (Mookhtiar *et al*., 1991; Gardner *et al*., 1997).

The pinky specificity loop (residues 740–769), which extends across the catalytic cleft and packs against the N-terminal domain, is another element that distinguishes RNAP from the pol I DNA polymerases (Figures 1 and 6). This segment contains N748, which has been shown to discriminate against similar promoter DNA sequences from phage T3 (Raskin *et al*., 1992), by forming contacts with promoter base pair –11. Location of N748 in the structure relative to the position of the active-site aspartate residues (D537, D812), which are expected to be positioned near promoter base-pair +1, unambiguously determine the polarity of template DNA in the nucleic acid cleft and the direction of RNA synthesis, which must proceed towards the N-terminal domain (Figure 6).

Perhaps the most striking difference between RNAP and the pol I DNA polymerases is the presence of the N-terminal domain, which is critical for establishment of the processive elongation stage of transcription. Its position in front of the polymerase domain closes off the cleft under the thumb sub-domain where primer template DNA is observed to bind in DNA polymerases (Eom *et al*., 1996; Doublie *et al*., 1998; Kiefer *et al*., 1998). This domain contributes to the unusually concave-shaped active-site cleft seen in RNAP and not in DNA polymerases (Brautigam and Steitz, 1998). Disruptions to this domain cause dramatic defects in the binding of promoter DNA/RNA and processive elongation (Ikeda and Richardson, 1986; Muller *et al*., 1988). Its orientation relative to the polymerase domain provides a solvent-accessible surface (helices F/G and the adjoining loop) that faces the active site and could interact with upstream promoter DNA or nascent RNA (Figure 6).

Model for interaction with template DNA and nascent RNA

The active-site cleft of RNAP has a bowl-like shape (Figure 6) that is strikingly different from that of other nucleic acid polymerases (Arnold *et al*., 1995; Brautigam and Steitz, 1998). Segments unique to RNAP create a highly concave active-site cleft, less open than those seen in DNA polymerase I (Eom *et al*., 1996), HIV reverse transcriptase (RT) (Kohlstaedt *et al*., 1992), or RB69 DNA

polymerase (Wang *et al*., 1997). Super-position of the active site of the *Taq* DNA polymerase–duplex DNA complex (Eom *et al*., 1996) onto RNAP, positions the 14/15mer duplex DNA within the active-site cleft of RNAP near regions implicated in catalysis (Bonner *et al*., 1992) and promoter recognition (Raskin *et al*., 1992). The unusual concave active-site cleft can only accommodate 6–8 bases of the modeled duplex which we presume occupies the position of the nascent RNA–DNA heteroduplex (Figure 6). The length that can be accommodated without clashes is in remarkable agreement with the length of RNA required for formation of a stable elongation complex (Martin *et al*., 1988). Furthermore, the clash with the N-terminal domain occurs with helices F and G, and the adjoining loop (residues 125–140), which in our structure are located above a series of residues whose mutation severely disrupts RNA binding and processivity (He *et al*., 1997). More precise discussion of the interaction of the N-terminal domain with DNA/RNA, along with conformational changes associated with various stages of the transcription cycle must await determination of the structures of appropriate complexes.

Sequence conservation among phage-type RNA polymerases

Comparison of the primary sequences of RNA polymerases from phage, mitochondria and chloroplasts in the context of the revised structure of RNAP, implies a conserved fold not only within the polymerase domain, but also in regions unique to pol I RNA polymerases (Figure 5). The significance of conserved residues in the polymerase domain that line the active site, many of which are conserved in DNA polymerases, RTs and telomerases, has been discussed (Delarue *et al*., 1990). Of particular interest is the striking sequence conservation among phage-type RNA polymerases in the β-strands (residues 724–740) that constitute a portion of the lysozyme binding site. These strands are conserved in structure though not in the amino acid sequence of pol I DNA polymerases (Figure 4C). Sequence and structural conservation of a major element of the lysozyme binding site leads us to wonder whether mitochondrial and chloroplast phage-type RNA polymerases, and perhaps even bacterial DNA polymerases, might also be regulated by binding at this site of a molecule analogous to T7 lysozyme.

Sequence similarities within structural elements unique to phage-type RNA polymerases strongly suggest that they will have a conserved structure. N-terminal domains of phage-type RNA polymerases vary greatly in sequence and length, a diversity partly accounted for by organellar targeting signals (Hedtke *et al*., 1997). Nonetheless, sequence similarity within the core of the N-terminal domain, spanning RNAP residues 140–325, which includes many buried side chains that presumably play a structural role, argues for the presence of this domain in phage-type RNA polymerases and a conservation of its function (Figure 5). The extended foot module also varies in length among the sequences we have examined. It culminates in a conserved C-terminal tetrapeptide containing hydrophobic or aromatic residues whose function, interaction with promoter DNA and nucleotides, is probably conserved (Gardner *et al*., 1997). Additionally, the length of the pinky specificity loop is conserved in phage-type RNA polymerases, but not its amino acid sequence. Two blocks of high sequence similarity in regions that flank the specificity loop attest to the accuracy of our alignment (Figure 5A). In bacteriophage T7, this loop contains determinants of specificity for T7 promoter–DNA sequences and discriminates against non-cognate promoters (Raskin *et al*., 1992). Homology within this element implies a conserved function in phage-type RNA polymerases. Insertions in the fingers sub-domains (590–620), whose function remains obscure, are only partially conserved amongst phage-type RNA polymerases (Figure 5A).

Discussion

In order to gain an insight into the mechanisms by which transcription can be regulated, we have studied the structure of T7 RNA polymerase bound to T7 lysozyme. Although it binds promoter DNA, RNA and nucleotides, the PL complex is arrested at the abortive initiation phase of the transcription cycle and loses the capability to synthesize long RNA products (Kumar and Patel, 1997; Zhang and Studier, 1997). Locking of RNAP in a non-processive conformation by interaction with LYS reduces initiation of RNA chains from phage promoters. This reduction in phage transcriptional capacity results in a biologically significant form of gene regulation during infection of *E.coli* whereby expression from the inefficient class II promoters is curtailed while that from the strong class III promoters continues until lysis (McAllister *et al*., 1981). The structure of the PL complex reveals that the lysozyme binding site is remote from the RNAP nucleic acid binding cleft, implying an indirect mechanism of inhibition. The binding site is composed of elements from the N-terminal domain and the palm and fingers sub-domains of RNAP. This observation leads to the hypothesis that lysozyme binding might prevent essential conformational changes that are expected to occur during the transition from abortive initiation to the stable elongation stage of transcription. This simple example serves as a structural paradigm for one way in which regulation of bacterial and eukaryotic transcription could be achieved.

T7 lysozyme is unchanged on binding to RNAP when compared with the uncomplexed structure (Cheng *et al*., 1994). The N-terminal five residues of LYS become structured and along with helix αA form the polymerase binding site. The surface of LYS that interacts with RNAP is in agreement with biochemical studies (Cheng *et al*., 1994). Furthermore, interaction with RNAP occludes the active site of the peptidase activity associated with T7 lysozyme. This finding is in agreement with functional studies that demonstrate polymerase inhibition and peptidase activity are mutually exclusive. Primary sequence comparisons with a homologous lysozyme from *Haemophilus influenzae* (Fleischmann *et al*., 1995) hint at a neat separation in the LYS structure of transcriptional inhibition and peptidase activities. Residues associated with transcription inhibition and polymerase binding cluster in the N-terminal 52 residues, while most residues involved with the peptidase activity map to the C-terminal 100 residues (Figure 5B).

The structure of T7 RNA polymerase obtained from this study differs substantially from that proposed by

Wang and co-workers (Sousa *et al*., 1993) with respect to assignment of the amino acid sequence to tertiary structure at some positions and chain orientation at others (Table II). A more accurate model of RNAP was obtained via 4-fold averaging of experimental electron density maps by domain, as well as by refinement of the structure, against data extending to 2.8 Å. The corrected structure has allowed re-interpretation of the structural basis of defects caused by site-directed mutations that affect interaction with LYS as well as binding of promoter DNA sequences and nascent RNA. Moreover, the revised structure will allow use of primary sequence alignments of RNA polymerases from bacteriophage, mitochondria from diverse eukaryotes and chloroplasts of some higher plants. An approximate model for nucleic acid binding to T7 RNA polymerase, constructed from sequence and structural alignments with the DNA polymerase I from *Thermus aquaticus*, yields insights into interactions with promoter DNA and nascent RNA.

Materials and methods

Sample preparation

Crystals of the T7 RNA polymerase–T7 lysozyme complex were prepared as described in Jeruzalmi and Steitz (1997). Isomorphous heavy-atom derivatives were prepared by soaking crystals in heavy-atom compounds dissolved in 10 mM MOPS, 5% PEG 8000, 20% sarcosine, 1 mM DTT, 0.02% NaN_3 and 10% propylene glycol. N7 platinated guanosine was prepared by reacting guanosine with an excess of $Pt[(NH_3)_3{\cdot}H_2O]$ (Reeder *et al*., 1996). Mercurated guanosine was prepared by reacting guanosine 5′-thiol (donated by Jennifer Doudna, Yale University) with $HgCl_2$.

Crystallographic structure determination

The structure of the T7 RNA polymerase–T7 lysozyme complex was determined from analysis of two crystal forms. Form I crystallized in space group C2 with cell parameters a = 320 Å, b = 93 Å, c = 229 Å, $\alpha = 90°$, $\beta = 129°$, $\gamma = 90°$ and with three complexes in the asymmetric unit. Crystals maintained at –175°C (Abdel-Meguid *et al*., 1996) diffracted to 3.3 Å resolution using synchrotron wiggler radiation (CHESS) and yielded data that merged with an R_{sym} ~8.2% in the 20–3.3 Å range. Crystal form III was in space group C2 with cell constants a = 270 Å, b = 93 Å, c = 63 Å, $\alpha = 90°$, $\beta = 103°$, $\gamma = 90°$ at T = –175°C with a single complex in the asymmetric unit. Diffraction data for crystal form III were recorded to 2.8 Å with an R_{sym} ~4.2% in the 30–2.8 Å range at CHESS (beamline A-1) with a charge coupled device detector.

Difference Patterson and difference Fourier maps were used to solve a series of isomorphous heavy-atom derivatives in each crystal form. Experimental phases were calculated with the MLPHARE program and modified with the constraints of solvent leveling and histogram matching as implemented in the program DM (CCP4, 1994). Phases in crystal form III confirmed (Read and Schierbeek, 1988) molecular replacement solutions obtained using a polyalanine model built from C_α atoms of T7 RNA polymerase (2RNP; Sousa *et al*., 1993) and the T7 lysozyme coordinates (1LBA; Cheng *et al*., 1994) as search models (Delano and Bilungen, 1995). Experimental phases in crystal form I were used to orient and position three PL complexes using a procedure to be described (D.Jeruzalmi and T.A.Steitz, unpublished).

Non-crystallographic symmetry (NCS) domain averaging between four complexes in two crystal forms was carried out using the RAVE (Kleywegt and Read, 1998), CCP4 (CCP4, 1994), and DEMON-ANGEL (Vellieux *et al*., 1995) software packages. Starting electron density maps were experimentally phased and contained no information from the model of Sousa *et al*. (1993) which was not used at any stage of the electron density calculations or averaging. Although molecular averaging was conducted with five transformations, significant improvements in the quality of the phases could only be obtained at moderate resolution (20.0–3.7 Å). Positional and temperature factor differences between the four complexes frustrated efforts to improve maps by NCS averaging, leading to a blurring of high resolution details. Nevertheless, the averaged maps clearly indicated deviations from the previously published model of RNAP.

Model building in crystal form III was performed using the interactive graphics program Ov5.10.3 (Jones *et al*., 1991) and checked against the three complexes in crystal form I. Chain connectivity was established unambiguously by inspection of the 4-fold averaged maps, while assignment of the amino acid sequence to tertiary structure was made using maps that were improved by the solvent-leveling and histogram-matching procedures in the DM program in crystal form III or with maps 2-fold averaged between crystal form III and complex #1 of crystal form I.

Model refinement

The PL model was refined using the CNS program (Brunger *et al*., 1998) against a data set extending to 2.8 Å collected from a crystal soaked in pHMBS (Form IIIa), not isomorphous to the 3.0 Å data set used in phasing and averaging. Rounds of positional refinement and restrained B-factor refinement were interspersed with manual re-building into $2F_o–F_c$ maps. These maps were calculated with F_cs derived either from averaging a family of 10 structures generated by multi-start torsion angle dynamics (L.M.Rice, Y.Shamoo and A.T.Brunger, submitted) or from a single structure where 2, 5, 10 or 50 residue segments had been omitted from the model prior to torsion angle dynamics (Figure 2). In these refinements, the maximum likelihood target (Pannu and Read, 1996; Adams *et al*., 1997) as implemented in CNS was found to be far superior to the crystallographic residual target. Partial model phases from either complete or omit models were combined with experimental phases using the program SIGMAA (CCP4, 1994) and subjected to a 5-fold averaging protocol (Form III, Form IIIa, Form I, #1, #2, #3) which generated maps clarifying portions of the N-terminal domain. Progress of refinement was monitored by reduction in R_{free} (Brunger, 1992). Model geometry was analyzed with the programs OOPS (Kleywegt and Jones, 1996) and PROCHECK (Laskowski *et al*., 1993). The final model consists of 923 residues with a Free*R* value of 31.8% with no outliers in the Ramachandran plot. Coordinates have been deposited with the Brookhaven Protein Data Bank under the accession code 1ARO.

Figure preparation

Figures were composed in programs BOBSCRIPTv1.0 (Esnouf, 1997), GRASPv1.25 (Nicholls *et al*., 1993) or RIBBONSv2.85 (Carson, 1991) with renderings performed in POVRAYv3.02 (http://www.povray.org), followed by processing in Macintosh CANVASv5.0. Figure 5A and B were generated using ALSCRIPT (Barton, 1993).

Acknowledgements

We thank Joseph Coleman for sharing unpublished data, Axel Brunger for unpublished programs and computing resources and John Kuriyan, also for computing resources. We gratefully acknowledge Paul Raccuia, Gerry Johnson and the CSB staff at Yale for excellent technical assistance, Kevin MacKenzie for introducing us to T7 lysozyme, Steve Bellon and Jennifer Doudna for assistance with nucleotide synthesis, Graham Cheetham for sharing unpublished data, and Elena Conti, Yousif Shamoo and Sarah Stallings for critical reading of the manuscript. We thank the Steitz laboratory for assistance in the collection of synchrotron X-ray diffraction data. This work was supported in part by NIH grant #GM-22778. Work at the Cornell High Energy Synchrotron Source and the MacCHESS facility are supported by the National Science Foundation (DMR-9311772) and the National Institutes of Health (RR-01646).

References

Abdel-Meguid,S.S., Jeruzalmi,D. and Sanderson,M.R. (1996) Preliminary characterization of crystals. In Jones,C., Mulloy,B. and Sandersen,M.R (eds), *Crystallographic Methods and Protocols*. Humana Press, Totowa, NJ, vol. 56, pp. 55–86.

Adams,P.D., Pannu,N.S., Read,R.J. and Brunger,A.T. (1997) Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.

Arnold,E., Ding,J., Hughes,S. and Hostomsky,Z. (1995) Structures of DNA and RNA polymerases and their interactions with nucleic acid substrates. *Curr. Opin. Struct. Biol.*, **5**, 27–38.

Barton,G.J. (1993) ALSCRIPT—A tool to format multiple sequence alignments. *Protein Eng.*, **6**, 37–40.

Bonner,G., Patra,D., Lafer,E. and Sousa,R. (1992) Mutations in T7 RNA polymerase that support the proposal for a common polymerase active site structure. *EMBO J.*, **11**, 3767–3775.

Bonner,G., Lafer,E.M. and Sousa,R. (1994a) Characterization of a set of T7 RNA polymerase active site mutants. *J. Biol. Chem.*, **269**, 25120–25128.

Bonner,G., Lafer,E.M. and Sousa,R. (1994b) The thumb subdomain of T7 RNA polymerase functions to stabilize the ternary complex during processive transcription. *J. Biol. Chem.*, **269**, 25129–25136.

Brautigam,C.A. and Steitz,T.A. (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.*, **8**, 54–63.

Brunger,A.T. (1992) The free R value: A novel statistical quality for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.

Brunger,A.T. *et al.* (1998) Crystallography and NMR system (CNS): A new software suite for macromolecular structure determination. *Acta Crystallogr.*, D in press.

Carson,M. (1991) Ribbons 2.0. *J. Appl. Crystallogr.*, **24**, 958–961.

CCP4 (1994) The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Cheng,X.D., Zhang,X., Pflugrath,J.W. and Studier,F.W. (1994) The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. *Proc. Natl Acad. Sci. USA*, **91**, 4034–4038.

Delano,W.L. and Brunger,A.T. (1995) The direct rotation function: Global correlation search applied to molecular replacement. *Acta Crystallogr.*, D in press.

Delarue,M., Poch,O., Tordo,N., Moras,D. and Argos,P. (1990) An attempt to unify the structure of polymerases. *Protein Eng.*, **3**, 461–467.

Doublie,S., Tabor,S., Long,A.M., Richardson,C.C. and Ellenberger,T. (1998) Crystal structure of a bacteriophage replication complex at 2.2 Å resolution. *Nature*, **391**, 251–257.

Eom,S.H., Wang,J. and Steitz,T.A. (1996) Structure of Taq polymerase with DNA at the polymerase active site. *Nature*, **382**, 278–281.

Erie,D.A., Yager,T.D. and von-Hippel,P.H. (1992) The single-nucleotide addition cycle in transcription: A biophysical and biochemical perspective. *Annu. Rev. Biophys. Biomol. Struct.*, **21**, 379–415.

Esnouf,R. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.*, **15**, 133–138.

Fleischmann,R.D. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Science*, **269**, 496–512.

Gao,G., Orlova,M., Georgiadis,M.M. and Hendrickson,W.A. (1997) Conferring RNA polymerase activity to a DNA polymerase: A single residue in reverse transcriptase controls substrate selection. *Proc. Natl Acad. Sci. USA*, **94**, 407–411.

Gardner,L.P., Mookhtiar,K.A. and Coleman,J.E. (1997) Initiation, elongation and processivity of carboxyl-terminal mutants of T7 RNA polymerase. *Biochemistry*, **36**, 2908–2918.

Gross,L., Chen,W.-J. and McAllister,W.T. (1992) Characterization of bacteriophage T7 RNA polymerase by linker insertion mutagenesis. *J. Mol. Biol.*, **228**, 488–505.

He,B., Rong,M., Durbin,R.K. and McAllister,W.T. (1997) A mutant T7 RNA polymerase that is defective in RNA binding and blocked in the early stages of transcription. *J. Mol. Biol.*, **265**, 275–288.

Hedtke,B., Borner,T. and Weihe,A. (1997) Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*. *Science*, **277**, 809–811.

Holm,L. and Sander,C. (1994) Searching protein structure databases has come of age. *Proteins*, **19**, 165–173.

Ikeda,R. and Bailey,P.A. (1992) Inhibition of T7 RNA polymerase by T7 lysozyme *in vitro*. *J. Biol. Chem.*, **267**, 20153–20158.

Ikeda,R.A. and Richardson,C.C. (1986) Interactions of the RNA polymerase of bacteriophage T7 with its promoter during binding and initiation of transcription. *Proc. Natl Acad. Sci. USA*, **83**, 3614–3618.

Ikeda,R.A. and Richardson,C.C. (1987) Enzymatic properties of a proteolytically nicked RNA polymerase of bacteriophage T7. *J. Biol. Chem.*, **262**, 3790–3799.

Inouye,M., Arnheim,N. and Sternglanz,R. (1973) Bacteriophage T7 lysozyme is an N-acetylmuramyl-L-alanine amidase. *J. Biol. Chem.*, **248**, 7247–7252.

Jacobo-Molina,A. *et al.* (1993) Crystal structure of Human Immunodeficiency Virus Type I reverse transcriptase complexed with double stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl Acad. Sci. USA*, **90**, 6320–6324.

Jeruzalmi,D. and Steitz,T.A. (1997) Use of organic cosmotropic solutes to crystallize flexible proteins: application to T7 RNA polymerase and its complex with the inhibitor T7 lysozyme. *J. Mol. Biol.*, **274**, 748–756.

Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **47**, 110–119.

Joyce,C.M. (1997) Choosing the right sugar: How polymerases select a nucleotide substrate. *Proc. Natl Acad. Sci. USA*, **94**, 1619–1622.

Joyce,C.M. and Steitz,T.A. (1994) Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.*, **63**, 777–822.

Kiefer,J.R., Mao,C., Braman,J.C. and Beese,L.S. (1998) Visualizing DNA replication in a catalytically active bacillus DNA polymerase crystal. *Nature*, **391**, 304–307.

Kim,Y., Eom,S.H., Wang,J., Lee,D.-S., Suh,S.W. and Steitz,T.A. (1995) Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature*, **376**, 612–616.

Kleywegt,G.J. and Jones,T.A. (1996) Efficient rebuilding of protein structures. *Acta Crystallogr.*, **D52**, 829–832.

Kleywegt,G.J. and Read,R.J. (1998) Not your average density. *Structure*, **5**, 1557–1569.

Kohlstaedt,L.A., Wang,J., Friedman,J.M., Rice,P.A. and Steitz,T.A. (1992) Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science*, **256**, 1783–1790.

Kostyuk,D.A., Dragan,S.M., Lyakhov,D.L., Rechinsky,V.O., Tunitskaya, V.L., Chernov,B.K. and Kochetkov,S.N. (1995) Mutants of T7 RNA polymerase that are able to synthesize both RNA and DNA. *FEBS Lett.*, **369**, 165–168.

Kumar,A. and Patel,S.S. (1997) Inhibition of T7 RNA polymerase: Transcription initiation and transition from initiation to elongation are inhibited by T7 lysozyme via a ternary complex with RNA polymerase and promoter DNA. *Biochemistry*, **36**, 13945–13962.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Ling,M.-l., Risman,S.S., Klement,J.F., McGraw,N. and McAllister,W.T. (1989) Abortive initiation by bacteriophage T3 and T7 RNA polymerases under conditions of limiting substrate. *Nucleic Acids Res.*, **17**, 1605–1619.

Livingstone,C.D. and Barton,G.J. (1993) Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput. Appl. Biosci*, **9**, 745–756.

Lyakhov,D.L., He,B., Zhang,X., Studier,F.W., Dunn,J.J. and McAllister,W.T. (1997) Mutant bacteriophage T7 RNA polymerase with altered termination properties. *J. Mol. Biol.*, **269**, 28–40.

Martin,C.T., Muller,D.K. and Coleman,J.E. (1988) Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry*, **27**, 3966–3974.

McAllister,W.T., Morris,C., Rosenberg,A.H. and Studier,F.W. (1981) Utilization of bacteriophage T7 late promoters in recombinant plasmids during infection. *J. Mol. Biol.*, **153**, 527–544.

McAllister,W.T. and Wu,H.-L. (1978) Regulation of transcription of the late genes of bacteriophage T7. *Proc. Natl Acad. Sci. USA*, **75**, 804–808.

Minnick,D.T., Astatke,M., Joyce,C.M. and Kunkel,T.A. (1996) A thumb sub-domain mutant of the large fragment of *Escherichia coli* DNA polymerase I with reduced DNA binding affinity, processivity and frameshift fidelity. *J. Biol. Chem.*, **271**, 24954–24961.

Moffat,B.A. and Studier,F.W. (1987) T7 Lysozyme inhibits transcription by T7 RNA polymerase. *Cell*, **49**, 221–227.

Mookhtiar,K.A., Peluso,P.S., Muller,D.K., Dunn,J.J. and Coleman,J.E. (1991) Processivity of T7 RNA polymerase requires the C-terminal Phe^{882}-Ala^{883}-COO^- or 'foot'. *Biochemistry*, **30**, 6305–6313.

Muller,D.K., Martin,C.T. and Coleman,J.E. (1988) Processivity of proteolytically modified forms of T7 RNA polymerase. *Biochemistry*, **27**, 5763–5771.

Nicholls,A., Bharadwaj,R. and Honig,B. (1993) GRASP: Graphical representation and analysis of surface properties. *Biophys. J.*, **64**, A166.

Ollis,D.L., Brick,P., Hamlin,R., Xuong,N.G. and Steitz,T.A. (1985) Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature*, **313**, 762–766.

Osumi-Davis,P.A., Aguilera,M.C.D., Woody,R.W. and Woody,A.Y.M. (1992) Asp 357, Asp 812 are essential and Lys 631, His 811 are catalytically significant in bacteriophage T7 RNA polymerase activity. *J. Mol. Biol.*, **226**, 37–45.

Pannu,N.S. and Read,R.J. (1996) Improved structure refinement through maximum likelihood. *Acta Crystallogr.*, **A52**, 659–668.

Patra,D., Lafer,E.M. and Sousa,R. (1992) Isolation and characterization of mutant bacteriophage T7 RNA polymerases. *J. Mol. Biol.*, **224**, 307–318.

Pelletier,H., Sawaya,M.R., Kumar,A., Wilson,S.H. and Kraut,J. (1994) Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer and ddCTP. *Science*, **264**, 1891–1903.

Raskin,C.A., Diaz,G., Joho,K. and McAllister,W.T. (1992) Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. *J. Mol. Biol.*, **228**, 506–515.

Read,R.J. and Schierbeek,A.J. (1988) A phased translation function. *J. Appl. Crystallogr.*, **21**, 490–495.

Rechinsky,V.O., Tunitskaya,V.L., Dragan,S.M., Kostyuk,D.A. and Kochetkov,S.N. (1993) Tyr-571 in involved is the T7 RNA polymerase binding to its promoter. *FEBS Lett.*, **320**, 9–12.

Rechinsky,V.O., Chernov,B.K., Dragan,S.M., Kostyuk,D.A., Tunitskaya, V.L. and Kochetkov,S.N. (1995) Targeted mutagenesis identifies ASP-569 as a catalytically critical residue in T7 RNA polymerase. *Mol. Gen. Genet.*, **247**, 110–113.

Reeder,F., Kozelka,J. and Chottard,J.C. (1996) *Inorg. Chem.*, **35**, 1413–1415.

Sastry,S.S. and Hearst,J.E. (1991) Studies on the interaction of T7 RNA polymerase with a DNA template containing a site specifically placed Psoralen cross-link I. characterization of elongation complexes. *J. Mol. Biol.*, **221**, 1091–1110.

Sousa,R., Patra,D. and Lafer,E.M. (1992) Model for the mechanism of bacteriophage T7 RNAP transcription initiation and termination. *J. Mol. Biol.*, **224**, 319–334.

Sousa,R., Chung,Y.J., Rose,J.P. and Wang,B.-C. (1993) Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature*, **364**, 593–599.

Sousa,R., Rose,J. and Wang,B.C. (1994) The Thumb's Knuckle. Flexibility in the thumb subdomain of T7 RNA polymerase is revealed by the structure of a chimeric T7/T3 RNA polymerase. *J. Mol. Biol.*, **244**, 6–12.

Steitz,T.A. (1993) DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.*, **3**, 31–38.

Steitz,T.A., Smerdon,S.J., Jaeger,J. and Joyce,C.M. (1994) A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science*, **266**, 2022–2025.

Studier,F.W. (1972) Bacteriophage T7. *Science*, **176**, 367–376.

Studier,F.W. and Dunn,J.J. (1983) Organization and expression of bacteriophage T7 DNA. *Cold Spring Harb. Symp. Quant. Biol.*, **47**, 999–1007.

Tjian,R. (1996) The biochemistry of transcription in eukaryotes: a paradigm for multisubunit regulatory complexes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **351**, 491–499.

Ujvari,A. and Martin,C.T. (1997) Identification of a minimal binding element within the T7 RNA polymerase promoter. *J. Mol. Biol.*, **273**, 775–781.

Vellieux,F.M.D.A.P., Hunt,J.F., Roy,S. and Read,R.J. (1995) DEMON/ ANGEL: A suite of programs to carry out density modification. *J. Appl. Crystallogr.*, **28**, 347–351.

von-Hippel,P.H., Bear,D.G., Morgan,W.D. and McSwiggen,J.A. (1984) Protein-nucleic acid interactions in transcription: A molecular analysis. *Annu. Rev. Biochem.*, **53**, 389–416.

Wang,J., Sattar,A.K.M., Wang,C.C. and Steitz,T.A. (1997) Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. *Cell*, **89**, 1087–1099.

Zhang,X. (1995) *T7 RNA Polymerase and T7 Lysozyme: Genetic, Biochemical and Structural Analysis of Their Interaction and Multiple Roles in T7 Infection*, State University of New York at Stony Brook, Stony Brook, NY, pp. 272.

Zhang,X. and Studier,F.W. (1997) Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme. *J. Mol. Biol.*, **269**, 10–27.

Received April 6, 1998; accepted May 1, 1998