Cell, Vol. 98, 811–824, September 17, 1999, Copyright ©1999 by Cell Press

Crystal Structure of *Thermus aquaticus* Core RNA Polymerase at 3.3 Å Resolution

Gongyi Zhang,*§ Elizabeth A. Campbell,* Leonid Minakhin,† Catherine Richter,* Konstantin Severinov,† and Seth A. Darst*‡
*The Rockefeller University
New York, New York 10021
†Waksman Institute and Department of Genetics
Rutgers, The State University of New Jersey
Piscataway, New Jersey 08854

Summary

The X-ray crystal structure of *Thermus aquaticus* core RNA polymerase reveals a "crab claw"–shaped molecule with a 27 Å wide internal channel. Located on the back wall of the channel is a Mg^{2+} ion required for catalytic activity, which is chelated by an absolutely conserved motif from all bacterial and eukaryotic cellular RNA polymerases. The structure places key functional sites, defined by mutational and cross-linking analysis, on the inner walls of the channel in close proximity to the active center Mg^{2+}. Further out from the catalytic center, structural features are found that may be involved in maintaining the melted transcription bubble, clamping onto the RNA product and/or DNA template to assure processivity, and delivering nucleotide substrates to the active center.

Introduction

RNA in all cellular organisms is synthesized by a complex molecular machine, the DNA-dependent RNA polymerase (RNAP). In its simplest bacterial form, the enzyme comprises at least four subunits with a total molecular mass of around 400 kDa. The eukaryotic enzymes comprise upward of a dozen subunits with a total molecular mass of around 500 kDa. The essential core component of the bacterial RNAP (subunit composition $\alpha_2\beta\beta'$) is evolutionarily conserved from bacteria to humans (Archambault and Friesen, 1993) (Figure 1). Sequence conservation points to structural and functional homologies, making the simpler bacterial RNAPs excellent model systems for understanding the multisubunit cellular RNAPs in general.

The basic elements of the transcription cycle were elucidated through study of the prokaryotic system. In this cycle, the RNAP, along with other factors, locates specific sequences called promoters within the double-stranded DNA, forms the open complex by melting a portion of the DNA surrounding the transcription start site, initiates the synthesis of an RNA chain, elongates the RNA chain completely processively while translocating itself and the melted transcription bubble along the DNA template, and finally releases itself and the completed transcript from the DNA when a specific termination signal is encountered. The current view is that the transcribing RNAP contains sites for binding the DNA template as well as forming and maintaining the transcription bubble, binding the RNA transcript, and binding the incoming nucleotide–triphosphate substrate.

Since the initial indications of DNA-dependent RNAP activity from a number of systems (Weiss and Gladstone, 1959; Huang et al., 1960; Hurwitz et al., 1960; Stevens, 1960) and the isolation of the RNAP enzyme from bacterial sources (Chamberlin and Berg, 1962), a wealth of biochemical, biophysical, and genetic information has accumulated on RNAP and its complexes with nucleic acids and accessory factors (von Hippel et al., 1984; Erie et al., 1992; Sentenac et al., 1992; Gross et al., 1996). Nevertheless, the enzyme itself, in terms of its structure–function relationship, remains a black box. An essential step toward understanding the mechanism of transcription and its regulation is to determine three-dimensional structures of RNAP and its complexes with DNA, RNA, and regulatory factors.

Low-resolution structures of bacterial and eukaryotic RNAPs, provided by electron crystallography, reveal a molecule shaped like a crab claw with a groove or channel that is an appropriate size for accomodating double-helical DNA (Darst et al., 1989, 1991, 1998a, 1998b; Schultz et al., 1993; Polyakov et al., 1995). To provide a more detailed framework for interpreting the existing genetic, biochemical, and biophysical information, as well as to guide further studies aimed at understanding the transcription process and its regulation, we have determined the three-dimensional structure of a bacterial core RNAP by X-ray crystallography at 3.3 Å resolution.

Results

Purification, Crystallization, and Structure Determination

The core RNAP isolated from *Thermus aquaticus* (see Experimental Procedures) comprised four distinct polypeptides. The three largest polypeptides were cloned and sequenced, identifying them as α, β, and β' (Figure 1 and Table 2). The fourth polypeptide (about 10.5 kDa) was tentatively identified as the ω subunit (see below). The isolated enzyme was active in a non–promoter-specific transcription assay (K. Murakami and S. A. D., unpublished observation).

Tetragonal crystals, space group $P4_12_12$ ($a = b = 201$, $c = 294$ Å), were grown by vapor diffusion (see Experimental Procedures). The crystals contained one 375.4 kDa core RNAP molecule per asymmetric unit, with a solvent content of 65%. Diffraction from the radiation-sensitive crystals was anisotropic, with reflections observed along the best and worst directions at 3.0 Å and 3.4 Å Bragg spacings, respectively. The structure was solved by the method of multiple isomorphous replacement (see Experimental Procedures and Table 1).

‡To whom correspondence should be addressed (e-mail: darst@rockvax.rockefeller.edu).
§Present address: National Jewish Medical and Research Center, 400 Jackson Street, Denver, Colorado 80206.

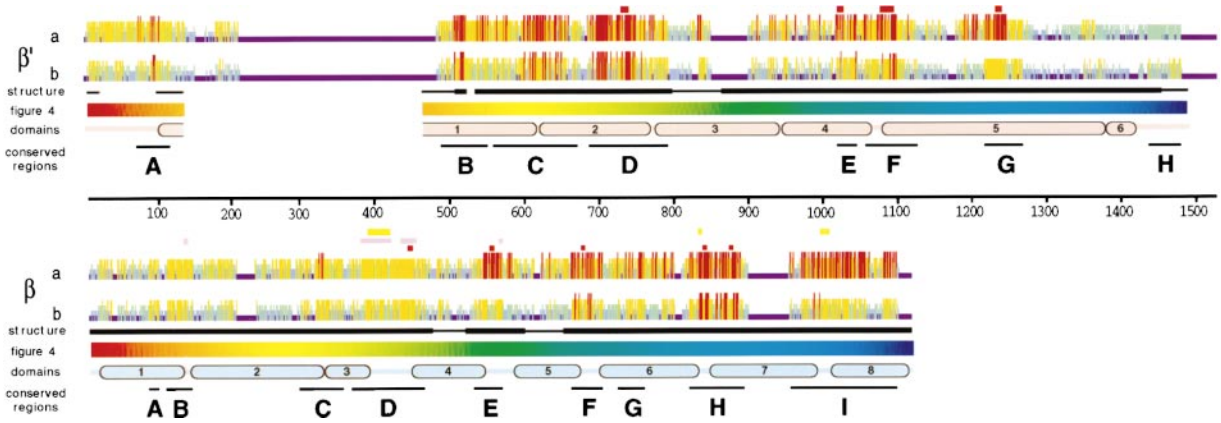

Figure 1. Sequence Features of *T. aquaticus* β and β′

The histograms represent the results of a sequence alignment of β′ (top) or β (bottom) homologs from 50 prokaryotic and chloroplast RNAP sequences (a) plus 26 eukaryotic and archaebacterial sequences ([b]; courtesy of V. Nikiforov). One hundred percent sequence conservation among all the sequences is represented by a tall red bar; less than 20% is represented by a small blue bar; intermediate levels are represented by orange, light green, and light blue bars. The numbered scale in the middle represents amino acid position. The first row underneath the histograms (labeled "structure") denotes the modeled portion of the structure, with full modeled structure and sequence represented by a thick black line and polyAla model represented by a thinner line. The next row underneath (labeled "figure 4") denotes the color coding used for the backbones of the large subunits in Figure 4, with the N terminus red, the C terminus blue, and a color gradient in between. The next row down (labeled "domains") denotes the domain architecture of the subunit. The structural domains are labeled 1–6 for β′ and 1–8 for β. The next row down (labeled "conserved regions") denotes the most highly conserved regions among all the prokaryotic, chloroplast, archaebacterial, and eukaryotic sequences, as initially identified for β′ by Jokerst et al. (1989) and for β by Sweetser et al. (1987) but with expanded conserved regions due to the larger number of aligned sequences. In addition, the original assignment of β′ conserved region E (Jokerst et al., 1989) is not supported by the more extensive alignment, and this region has been redefined. Above the histograms, red bars denote the positions of cleavage sites by hydroxyl-radicals generated from the active center metal-chelation site (Zaychikov et al., 1996; Mustaev et al., 1997), magenta bars represent the locations of mutations that confer rifampicin resistance (Jin and Gross, 1988; Severinov et al., 1994), and yellow bars denote the positions of cross-links to initiating nucleotide analogs (Mustaev et al., 1991; Severinov et al., 1995; Zaychikov et al., 1996).

Modeling and Refinement

The initial MIR map showed protein-solvent boundaries and contained some identifiable α helices. Density modification resulted in a dramatically improved map (Figure 2). The fold of the previously solved *E. coli* α subunit N-terminal domain (NTD) dimer (Zhang and Darst, 1998) was easily recognized, and the αNTD structure was modeled and refined as described in Experimental

Table 1. Crystallographic Data

Data Set	Resolution (Å)	R_{merge}[a] (%) (+/− ano)	No. of Unique Reflections (+/− ano)	Total Observations[b]	Completeness[c] (%) (+/− ano)	Phasing Power[d] (Centric/Acentric/Ano)	No. of Sites
TriMetPb[e]	3.2	10.0/7.6	78,640/122,450	181,691	82.5/67.4	−/−/0.22	1
$HgCl_2$[e]	3.2	6.2/5.5	70,627/103,250	245,952	77.1/50.5	1.10/1.23/0.58	4
EthylHgCl[e]	3.2	7.5/5.7	77,123/120,759	188,044	78.5/64.5	1.10/1.45/0.48	3
Ta_6Br_{14}[e]	3.7	7.7/5.8	47,995/72,245	107,255	75.4/59.9	0.55/0.68/0.50	5
Mersalyl[f]	3.2	7.7/6.2	90,039/169,685	248,398	92.5/75.3	1.06/1.48/0.75	1
$KAu(CN)_2$[f]	3.3	12.1/9.8	70,851/104,869	145,942	79.8/62.1	0.39/0.44/0.12	1
EMTS[f]	3.2	11.3/8.7	77,657/111,598	140,480	79.4/59.8	1.22/1.64/0.81	5
Ir_4[g]	3.0	9.5/8.0	866,353/161,702	308,116	74.7/71.4	0.46/0.65/0.32	4
$HgCl_2$[g]	3.0	9.1/8.0	926,452/15,026	346,661	77.0/65.4	1.17/1.50/0.98	4
Se Met[e,h]	4.0	9.7	45,071	133,158	80.1		45

Mean Figure of Merit[d]

Resolution (Å)	40.36–8.27	40.36–5.91	40.36–4.84	40.36–4.20	40.36–3.76	40.36–3.44	40.36–3.18	Overall
No. of reflections	6,070	10,153	12,928	15,112	17,054	18,761	20,004	100,082
F.O.M.	0.756	0.674	0.543	0.406	0.278	0.165	0.085	0.341

[a] $R_{merge} = \Sigma|I_j - \langle I \rangle|/\Sigma I_j$, with or without Bijvoet pairs treated as equivalent.
[b] Total observations, the number of full and partial observations measured with nonnegative intensity to the indicated resolution.
[c] Completeness, the percentage of possible unique reflections measured with $I/\sigma(I) \geq 0$ to the indicated resolution.
[d] Phasing power and figure of merit are from SHARP (de La Fortelle et al., 1997).
[e] Data set was collected at CHESS A1, F1, or F2.
[f] Data set was collected at APS 14-BM-C.
[g] Data set was calculated at NSLS X25.
[h] SeMet data was not included in phase calculations.

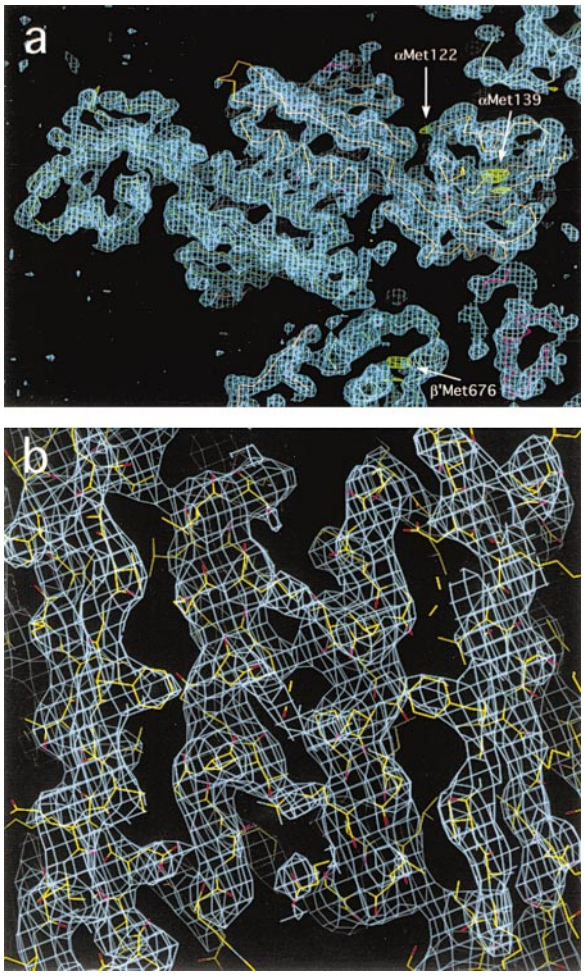

Figure 2. Experimental Electron Density Maps

(a) A thin slice from the MIR electron density map, after density modification (see Experimental Procedures), showing the region corresponding to the αNTD dimer. The α carbon backbone of one αNTD monomer from the final, refined structure is shown in yellow, the other in green. At the upper right is a small portion of a symmetry-related molecule. At the bottom are some regions from other subunits of the RNAP. Shown in green are selenomethionine difference Fourier peaks contoured at 3σ. The corresponding methionine residues in the structure are labeled.

(b) A close-up view of the MIR electron density map, after density modification, in the region of the αNTD dimer interface, with the refined model superimposed.

Procedures (Figure 2). Phase combination and multidomain noncrystallographic symmetry averaging were then used to obtain a slightly improved map. This map was exceptionally clean, with secondary structural elements and well-connected main chain density over most of the structure, allowing building of a polyalanine model containing about 85% of the expected number of residues for β and β′. Side chain density, while present in much of the map, was weak to nonexistent in other regions. For this reason, selenomethionyl core RNAP was prepared and crystallized (see Experimental Procedures). The resulting Fourier-difference peaks aided in the localization of methionine residues during modeling (Figure 2a).

After positional refinement of an initial model, the resulting phase-combined maps revealed additional side chain density, allowing adjustment of the model and assignment of additional sequence. The current model (Table 2) contains about 70% of the main chain of β′, the complete main chain of β (except for a few residues at each terminus), the αNTD dimer, a 91-residue polyAla model of ω, one Mg^{2+} ion (chelated at the active center), and one Zn^{2+} ion. Lacking electron density and presumably disordered in the crystal are both α C-terminal domains, as well as a 74-residue segment of β′ that includes a Zn^{2+}-binding motif along with most of β′ conserved region A (β'_A). The region of primarily helical, well-defined electron density assigned to ω was completely detached from any other density and was at odds with the secondary structure predicted for β'_A (the only region not assigned that was large enough to account for the density), which was completely β sheet (Rost and Sander, 1993). Secondary structure predictions using the sequence of either *E. coli* or *Deinococcus radiodurans* (evolutionarily closely related to *T. aquaticus*) ω matched the structure of this portion almost to the residue, leading to its assignment as ω. A nonconserved sequence of 330 residues inserted between β'_A and β'_B (Figure 1) is currently not modeled. Electron density for this domain is present but weak and generally not well connected. Several stretches of residues are modeled as polyAla. These include linker regions between β'_D and β'_E, as well as between $\beta_{D\text{-}E}$ and $\beta_{E\text{-}F}$. The R factor for the current model is 0.329 for data from 8–3.3 Å resolution ($R_{free} = 0.399$). The model presented here cannot be characterized as well refined; on the basis of this refinement alone, we cannot rule out local areas of misassigned sequences (sequence register shifts) in some regions of the structure. Nevertheless, with the exception of β'_A, which is disordered and not modeled, the conserved regions of the large subunits are generally well defined. In addition, the structure of the αNTD dimer was easily built from the known structure with no ambiguities. Serving to limit the possibility of errors in the structure was the availability of selenomethionine difference peaks; within the modeled portion of the structure, there are 42 Met residues, 38 of which correlate with selenomethionine peaks. Also, within the modeled portion of the structure, there are 9 Cys residues; 7 of these are bound by various Hg derivatives (Table 1), and the other two are buried and are not solvent accessible. Furthermore, the binding site for a single-site Pb derivative was interpreted to be the known site of Mg^{2+} chelation in the enzyme active center by 3 Asp residues in the absolutely conserved -NADFDGD- motif of β'_D (Zaychikov et al., 1996). It was subsequently shown that Pb ions bind to this site in the protein with very high affinity (A. Mustaev, personal communication). To resolve the structure more completely, further refinement is in progress, as well as collection of more accurate and higher-resolution data.

Further support for the structure comes from the fact that it explains a wide range of independent biochemical, biophysical, and genetic data available in the literature (which was not used to guide the model-building process). The available evidence supporting the model includes the following. (1) With only a few exceptions, the structure corresponds to the predicted secondary

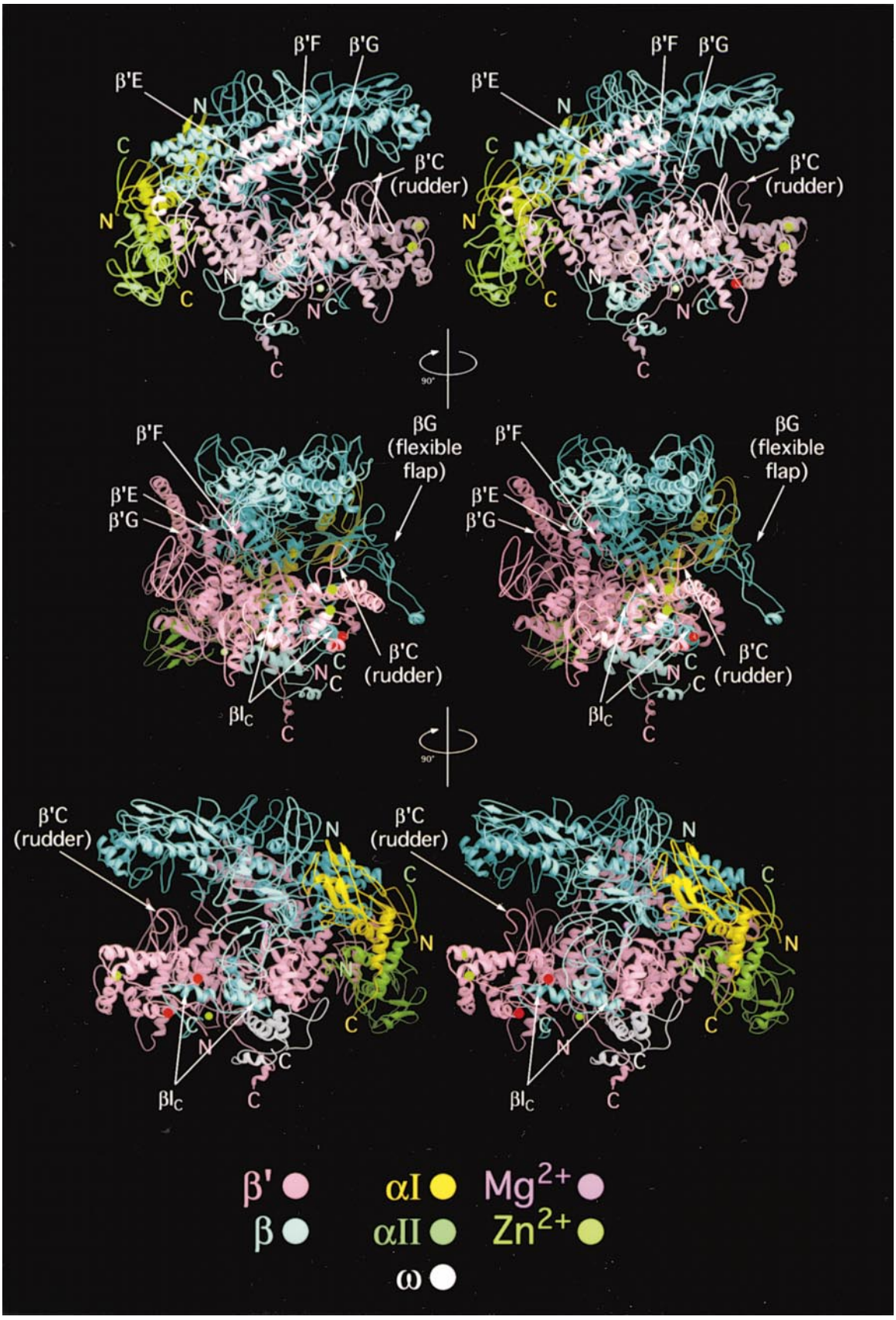
β'F
β'G
β'E
N
C
β'C
(rudder)
90°
βG
(flexible
flap)
βI_C
β'
β
αI
αII
ω
Mg^{2+}
Zn^{2+}

Table 2. Structural Model

Subunit	n[a]	M_r (kDa)	No. of Residues		Regions Modeled
			Sequence	Model	
β′	1	170.7	1,525	1,077	4–22, 96–132, 462–523, 535–1493 (polyAla: 4–22, 96–132, 462–509, 797–869, 1451–1493)
β	1	124.4	1,119	1,112	2–1113 (polyAla: 478–521, 599–652)
α	2	34.9	313	226	6–231
ω	1	10.5	91	91	1–91 (polyAla: 1–91)
Total	5	375.4	3,361	2,732	

[a] Number of copies of the subunit in the RNAP assembly.

structure (Rost and Sander, 1993), which is expected to be accurate because of the large number of sequences used in the prediction for β and β′ (50 and 86, respectively). (2) Prokaryotic RNAPs are inhibited by the antibiotic rifampicin, which binds with high affinity to a genetically well-characterized site in β. Mutations conferring rifampicin resistance are scattered throughout the primary sequence of β (Jin and Gross, 1988; Severinov et al., 1993) (Figure 1) but are clustered together in the structure (Figure 5, bottom panel). (3) Genetic and cross-linking studies have identified residues in widely separated regions of the β subunit sequence that are directly involved in binding the initiating NTP substrate or are within a few angstroms of the site (Mustaev et al., 1991; Severinov et al., 1995) (Figure 1), but these residues are clustered together in the structure (Figure 5, bottom panel). (4) A fusion between the C terminus of β and the N terminus of β′ in *E. coli* shows no detectable defects in vivo or in vitro (Severinov et al., 1997), and the fusion occurs naturally in some bacterial species (Zakharova et al., 1999). These two sites are immediately adjacent to one another in the structure (Figure 3). (5) Known sites of protease sensitivity in intact RNAP (Borukhov et al., 1991; Severinov et al., 1992) are exposed on the surface of the structure. (6) Mustaev et al. (1997) used Fe^{2+}-generated hydroxyl-radical cleavage to identify nine sites widely separated in the primary sequences of β (five sites) and β′ (four sites; Figure 1) that are all close to the active center Mg^{2+}. These sites are all less than 20 Å from the active center Mg^{2+} in the structure (Figure 5, top panel). (7) Mustaev et al. (1994) used chimeric rifampicin–ATP compounds to show that the rifampicin-binding site and the initiating NTP substrate site (the i site) are within 15 Å of each other, which is explained by the structure (Figure 5, bottom panel).

General Architecture

The shape and size of the *T. aquaticus* core RNAP X-ray structure (Figures 3, 4, and 6) correspond extremely well with the low-resolution structure of *E. coli* core RNAP from cryoelectron microscopy (Darst et al., 1998a, 1998b). The shape is reminiscent of a crab claw, with an internal groove or channel running along the full length (between the claws). One arm of the claw is primarily the β subunit, the other primarly β′. The molecule is about 150 Å long (from the back to the tips of the claws), 115 Å tall, and 110 Å wide (parallel with the channel). The channel has many internal features, but the overall width is about 27 Å.

The overall architecture of the RNAP, as well as the individual subunits and their organization and interactions, can be viewed in Figure 3. Specific regions of the structure discussed in the text are also labeled in Figure 3, but because of the size and complexity of the structure, it is difficult to follow any particular polypeptide chain. The overall paths of the polypeptide backbones of the β and β′ subunits are illustrated in Figure 4 using a color gradient from red (N terminus) to blue (C terminus). Specific regions of each subunit can be roughly located by referring to the same color code illustrated below each primary structure schematic in Figure 1.

Subunit Interactions

A detailed analysis of RNAP intersubunit contacts will be presented elsewhere. Some points of general interest are summarized here. As expected, the RNAP subunits make extensive interfaces with each other. β and β′ each contribute about 17% of their solvent-accessible surface (Lee and Richards, 1971) to contacts with other subunits. Indicative of its presumed role in assembly, each αNTD monomer contributes about 24% of its solvent-accessible surface to intersubunit contacts. The structure supports the view that the αNTD dimer functions to aid the assembly of β and β′ but does not participate directly in catalysis. In fact, no residues of the αNTD dimer have access to the internal channel of RNAP where catalysis takes place.

β regions F, G, H, and I contact the αNTD dimer almost

Figure 3. Structure of *T. aquaticus* Core RNAP

Stereo-RIBBONS diagrams of the three-dimensional structure of core RNAP. Various features discussed in the text are labeled. The break in the chain of β′ due to the disordered region A is indicated by red dots (the gap corresponds to residues 23–95). The break in the chain of β′ due to the unmodeled, nonconserved region is indicated by green dots. (The gap corresponds to residues 133–461.) The individual subunits are color-coded as indicated at the bottom of the figure. The Mg^{2+} ion chelated at the active center is indicated by a magenta sphere. The Zn^{2+} ion bound in β′ (see text) is indicated by a light green sphere. (Top) View roughly parallel with the main axis of the RNAP channel. (Middle) Top view rotated 90° clockwise about the vertical axis. (Bottom) Top view rotated 180° about the vertical axis, giving a view down the opposite end of the main channel.

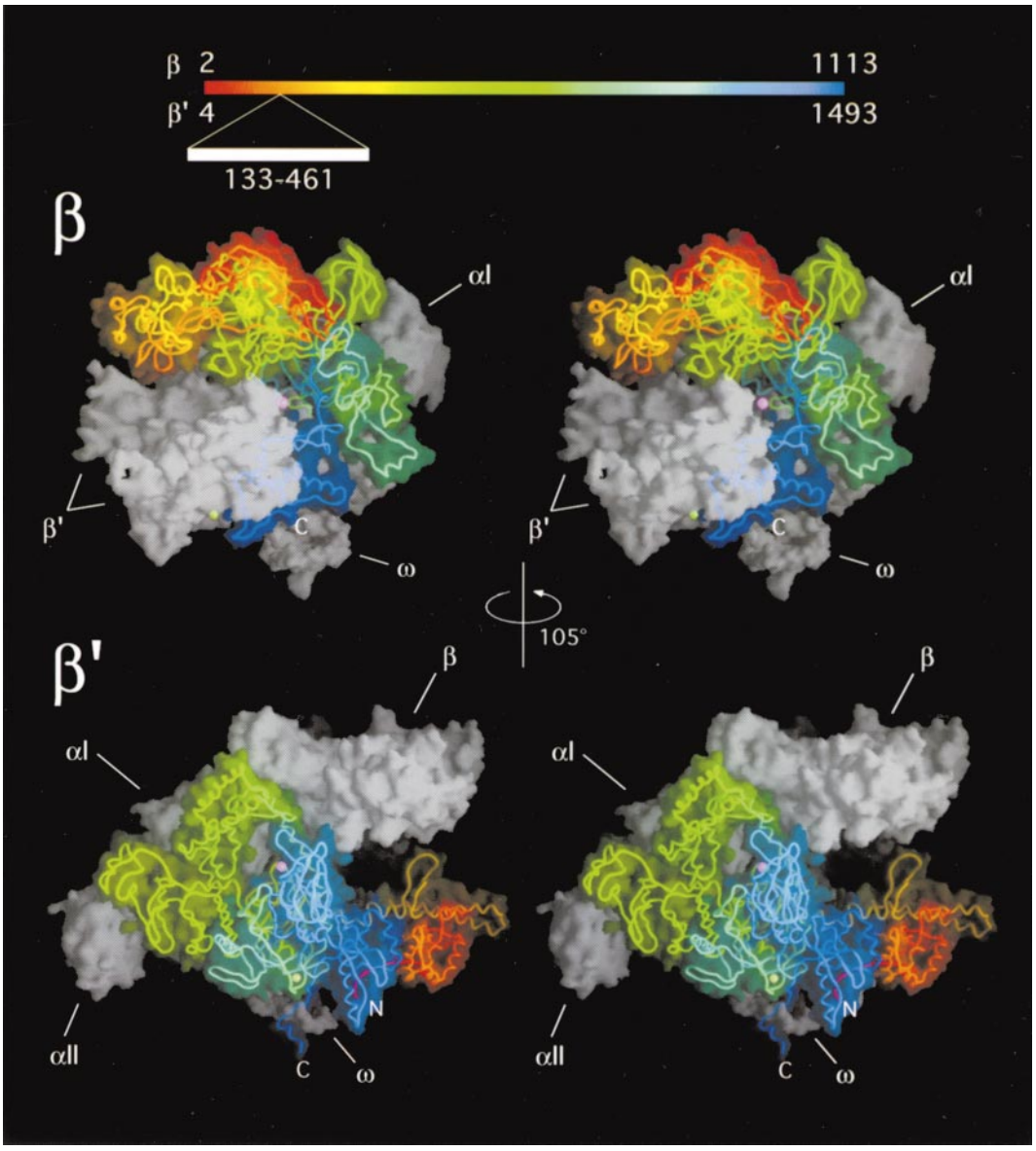

Figure 4. The RNAP β and β′ Subunits

Two stereo views of the RNAP structure (represented by transparent molecular surfaces), displayed using the program GRASP (Nicholls et al., 1991). The paths of the polypeptide backbone for β (top) or β′ (bottom) are shown as worms and color-coded with a gradient from the N to C terminus according to the scheme shown at the top of the figure (the color coding is also shown in Figure 1 with closer reference to the features of the primary structure). The nonconserved domain of β′ (residues 133–461) is not included in the color gradient. The surfaces are colored according to the worm color, or else are white for the other RNAP subunits. The Mg^{2+} ion chelated at the active center is indicated by a magenta sphere. The Zn^{2+} ion bound in β′ (see text) is indicated by a light green sphere.

exclusively through only one of the αNTD monomers (denoted αI; Figure 3), with the primary interface being β_H, consistent with the findings of Wang et al. (1997). β′ regions C, D, G, and H contact exclusively the other α monomer (αII). The interactions that β and β′ make with the αNTD dimer closely match the hydroxyl-radical protein footprinting data of Heyduk et al. (1996).

β and β′ make extensive interactions with each other. A major interface between the two large subunits occurs at the base of the channel where the active center Mg^{2+} is chelated (Figure 3). Particularly critical are interactions between β regions H and I and β′ region D, which position the -NADFDGD- motif of β'_D for chelating the active site Mg^{2+} (Figure 5, bottom panel).

Of particular importance is β_I, which makes a number of critical interactions. An N-terminal part of β_I (residues 974–979) makes contacts with αI that are critical for the formation of the $\alpha_2\beta$ assembly intermediate (Wang et al., 1997). The middle of β_I (residues 998–1008) contacts β′ regions C, G, and H along with β_H and β'_D to help form the catalytic center (Figure 5, bottom panel). Finally, the C-terminal part of β_I (residues 1009–1099), which is

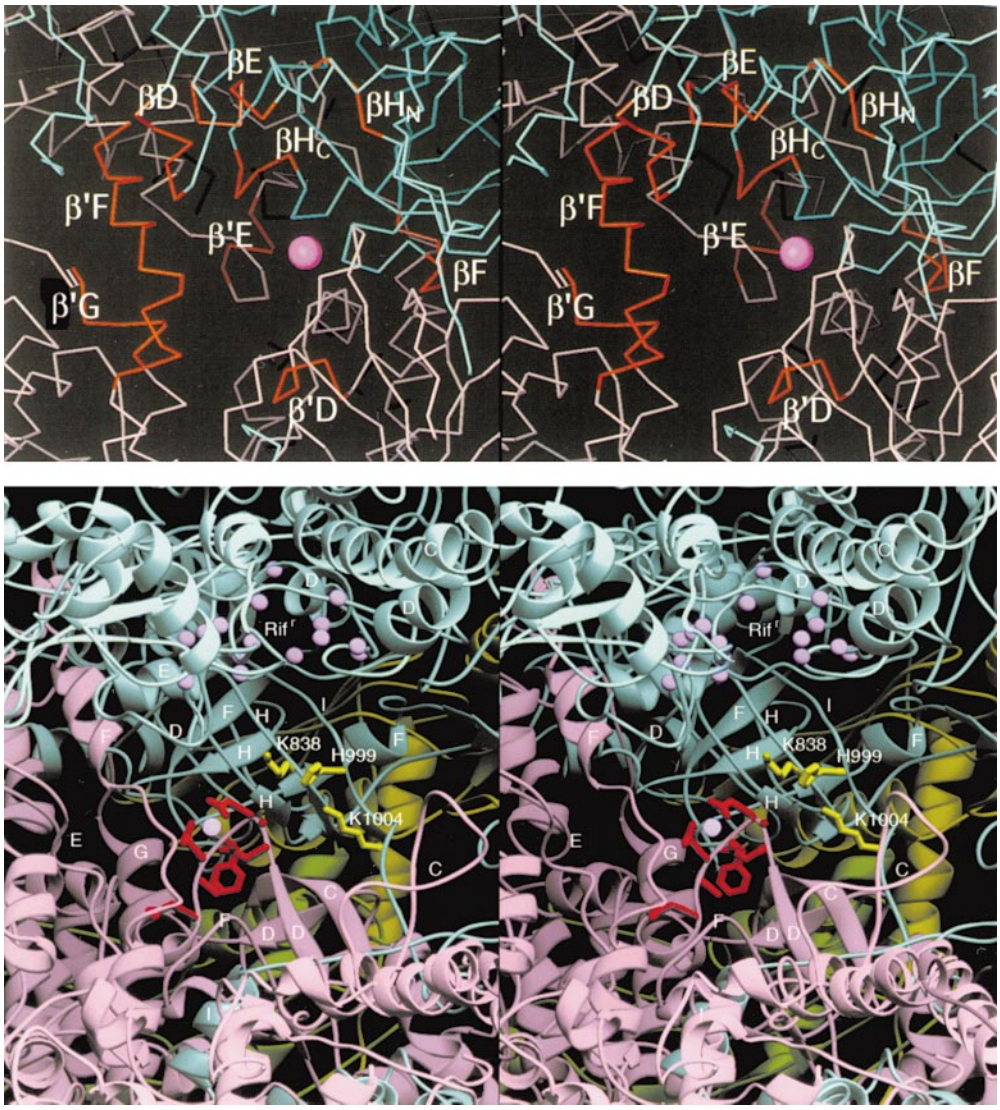

Figure 5. Active Center of RNAP

(Top) Stereo view showing the α carbon backbone of β (cyan) and β′ (rose) around the region of the active center Mg^{2+} ion (magenta sphere). The sites of hydroxyl-radical cleavage of the β and β′ polypeptides by Fe^{2+} substituted in the active center Mg^{2+} site (Zaychikov et al., 1996; Mustaev et al., 1997) are colored red and labeled according to the subunit and the conserved region. (All of the cleavage sites occur within conserved regions of the large subunits.)

(Bottom) Stereo RIBBONS diagram of the RNAP active center. The view is roughly the same as Figure 3B. The RNAP subunits are color-coded as in Figure 3 (β′, rose; β, cyan; αI, yellow). The locations of some conserved regions of β and β′ are labeled with white letters. The active center Mg^{2+} is shown as a pink sphere. The side chains of the absolutely conserved -NADFDGD- motif from β'_D, responsible for chelating the active center Mg^{2+} (Zaychikov et al., 1996), are shown in red. The side chains shown in yellow are three residues from β that have been mapped to within a few angstroms of the initiating NTP (βK838, βH999, and βK1004). The magenta spheres denote the α carbons of amino acid positions where substitutions confer rifampicin resistance (Rif^r). These residues line a small pocket on the roof of the main RNAP channel.

required to recruit β′ into the $\alpha_2\beta$ assembly intermediate (Wang et al., 1997), forms a separate domain from the rest of β but is almost completely surrounded by β′ regions B, C, D, and H (Figures 3 and 4). Overall, β regions A, B, and C are the only conserved regions of the two large subunits that do not make intersubunit contacts.

The ω subunit makes contacts only with β′, consistent with the cross-linking results of Gentry and Burgess (1993). The ω subunit completely wraps around the C-terminal tail of β′ (Figures 3 and 4), suggesting ω may play a chaperonin role in the final stages of RNAP assembly (Mukherjee and Chatterji, 1997).

Subunit Structure

The structure of the αNTD dimer in the core RNAP is almost identical to the isolated αNTD structure (Zhang and Darst, 1998) except for domain movements. In the RNAP structure, domain II of each αNTD monomer is rotated toward the RNAP. Domain II (along with domain

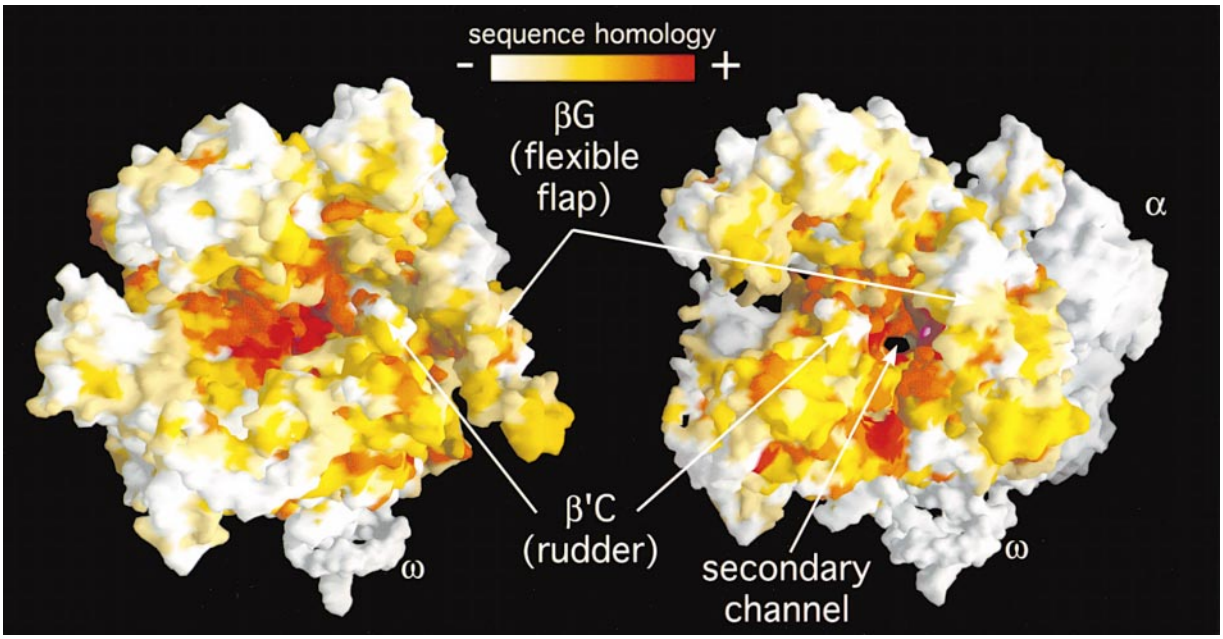

Figure 6. Sequence Conservation in β and β′ Mapped onto the Core RNAP Structure

Molecular surface representations of core RNAP color-coded according to amino acid sequence conservation within the β and β′ subunits (the α and ω surfaces are not color-coded and are white), with low sequence conservation shown as white, very high (100%) shown in red, with a gradient in between. Some structural features discussed in the text are labeled. Displayed using the program GRASP (Nicholls et al., 1991). (Left) The same view as shown in Figure 3B. (Right) A similar view as shown in Figure 3C. The active site Mg^{2+} is visible as a magenta sphere. In this view, one can see directly through the molecule down the axis of the secondary channel.

I) of each αNTD monomer makes interactions with β (αI) or β′ (αII), but the interactions between domain II of αI and β are much more extensive. In fact, domain II of αII makes contacts with a region of β′ between regions D and E that shows little sequence conservation with eukaryotes. Thus, we predict that this interaction is not critical for RNAP assembly.

As expected for such large proteins, both β and β′ comprise a number of relatively distinct domains (Figure 1). Two domains of β are unique in that they extend away from the main body of β and do not interact with other β domains. One of these, the already mentioned C-terminal part of β_I (β structural domain 8), makes extensive interactions with β′. The second domain, which spans β_F and β_H and includes β_G (β domain 6), forms a flap-like structure that appears to be flexibly connected to the rest of β (Figure 3); the position of this domain in the crystal structure is fixed only by crystal contacts with symmetry-related RNAP molecules.

The overall topology of β′ is circular in that the N and C termini are near each other (Figures 3 and 4). A domain of β′ that includes β'_E extends up and interacts with β on the face of the RNAP molecule nearest the viewer in Figure 3. Region F of β′ is most remarkable; it begins in the upper domain of β′ where β'_E ends, then forms a helical segment and loop that traverses directly across the middle of the main channel (Figure 3), then ends firmly anchored in the main body of β′. The active center Mg^{2+} is positioned at the base of the main channel directly across from the β'_F helix (Figure 7a). The β'_F helix conspires with β'_G, which forms a long loop that extends into the main channel, to form a wall-like structure that bifurcates the main channel into two separate channels (Figure 7a). The secondary channel thus formed is rougly 10–12 Å in diameter, not large enough to accommodate double-stranded nucleic acid (either DNA–DNA or DNA–RNA). Furthermore, examination of the structure suggests that threading of a single strand of DNA (such as in the melted region of the transcription bubble) through the secondary channel is unlikely. To achieve this without breaking a covalent bond in the DNA, the secondary channel would have to be opened by disrupting the extensive interactions between β′ regions E and F with β at the N-terminal end of the β'_F helix. Finally, a coiled coil–like structure extends from the main channel and supports another loop-like structure that protrudes upward, forming a rudder-like feature comprising β′ region C (labeled β'_C rudder in Figure 3). This feature can be seen clearly in the color-coded backbone of β′ as the yellow coiled coil and loop extending from the right side of the main channel in Figure 4.

β′ contains an unusual Zn^{2+}-binding motif (Figure 3) comprising four Cys residues between regions F and G (Markov et al., 1999). Three of the Cys residues are arranged in a sequence reminiscent of a Zn^{2+}-binding motif (β′1195-**C**X_6**C**X_2**C**). The fourth Cys participating in the Zn^{2+} chelation is β′Cys1113, 82 residues away, explaining why this was not identified as a Zn^{2+}-binding site from sequence analysis. The four Cys residues are absolutely conserved in prokaryotes (they correspond to positions 814, 888, 895, and 898 of *E. coli* β′) but are not conserved in eukaryotes. This, along with its location on the bottom side of β′ on the outside of the channel (Figure 3), suggests it plays a critical structural role in

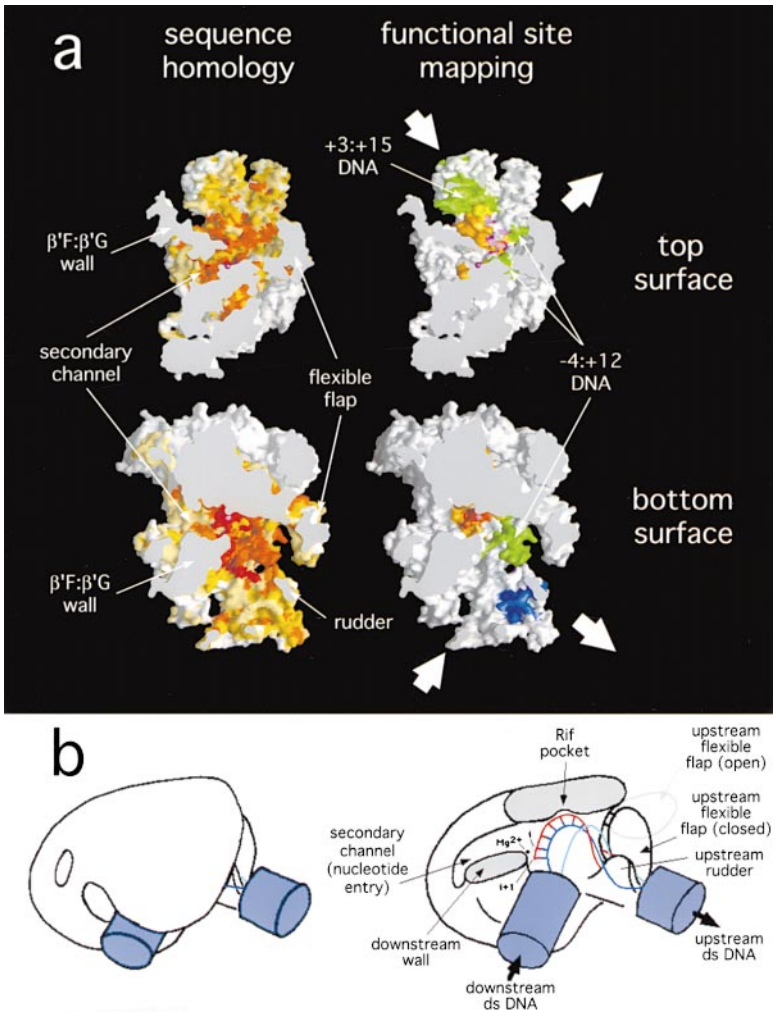

Figure 7. RNAP Structure–Function Relationship

(a) Molecular surface representations of the "open book" views of the inside of the RNAP channel. The top row shows the inside, top surface of the channel (primarily β), and the bottom row shows the inside, bottom surface (primarily β′). Colored gray are the parts of the protein structure that have been sliced away. (The gray surfaces of the top and bottom views do not match because the slicing and viewing angles are different to afford the best views of the structural features discussed.) The active center Mg^{2+} is visible as a magenta sphere. On the left, the sequence conservation is mapped onto the structure as in Figure 6. On the right, various functional sites determined from DNA and RNA cross-linking experiments are mapped onto the structure. The color coding is as follows: red, absolutely conserved -NADFDGD- motif of β'_D; orange, cross-links to various probes positioned at the 3′-end of the RNA transcript (Markovtsov et al., 1996; Nudler et al., 1998); yellow, cross-links to various probes position at the 5′-end of the i site NTP substrate (Mustaev et al., 1991; Severinov et al., 1995; Zaychikov et al., 1996); green, cross-links from probes incorporated into specific positions of the template strand of the DNA (Nudler et al., 1996); blue, a cross-link mapped from a probe incorporated at the −10 position of the RNA transcript (Nudler et al., 1998).

(b) Schematic model of the structure of a ternary transcription complex. Double-stranded DNA is represented as blue cylinders. The DNA template strand is shown as a blue line; the nontemplate strand, a cyan line; the RNA transcript, a red line. Very little information is available to position the nontemplate DNA strand within the model; it is shown here for illustrative purposes only. (Left) View with intact RNAP molecule. (Bottom) Same view but with parts of the RNAP cut away (shown in gray) to reveal the inner workings of the complex, which are labeled.

the folding of β′ that is performed by some other subunit of the eukaryotic enzymes.

Active Center

As expected, the three Asp residues within the absolutely conserved -NADFDGD- motif of β'_D chelate a Mg^{2+} ion (Figure 5). Substitution of these Asp residues by Ala gives rise to a dominant-lethal phenotype, which is explained by the in vitro ability of the mutant RNAP to occupy promoter sites on the DNA and form stable open complexes that lack any detectable catalytic activity (Zaychikov et al., 1996), identifying the chelated Mg^{2+} as the catalytic center of the enzyme. The hydroxyl-radical cleavage experiment of Mustaev et al. (1997) identified nine sites widely separated in the primary sequences of β (five sites) and β′ (four sites; Figure 1) that must be close to the active center Mg^{2+}. All of the mapped hydroxyl-radical cleavage sites converge near the active center Mg^{2+} in the structure (Figure 5, top panel). One cleavage site (β′E in Figure 5, top panel) is 20 Å from the active center Mg^{2+}, and the others are 12 Å or less, although some residues adjacent to the mapped regions (never more than 3 or 4 residues) are sometimes even closer to the active center Mg^{2+} but were not mapped. Two additional protein fragments are within 12 Å of the active center Mg^{2+}. One is centered about βHis999 (within β_I, see Figure 5, bottom panel). This fragment would not have been mapped in the hydroxyl-radical cleavage experiment because it lies C-terminal of the site used to radioactively label the β subunit. A second region is centered about β′ residue 632 (within β'_C). The hydroxyl-radical cleavage sites were mapped by analysis of the electrophoretic mobility of the protein cleavage products. Small errors in the analysis due to anomalous mobilities of the protein fragments could easily account for the small discrepancies with the structure noted above. Furthermore, the hydroxyl-radical cleavage experiment was done on a binary complex of RNAP holoenzyme with promoter DNA. Conformational changes of the RNAP around the active center, as well as protection of some protein fragments from hydroxyl-radical cleavage by the presence of the DNA or the σ subunit, could also account for these discrepancies.

The high degree of sequence conservation between the large RNAP subunits from prokaryotes to eukaryotes (Figure 1) points to structural homologies, which are borne out by low-resolution structures from electron microscopy (Darst et al., 1991, 1998a, 1998b). The multiple functions performed by the elongating RNAP, which result in the rapid, highly processive, and accurate synthesis of RNA complementary to the template strand of the DNA, likely leave little room for evolutionary variability in the region surrounding the active center. This is indeed the case (Figures 6 and 7a); regions of very high sequence conservation (approaching 100%) are concentrated around the active center Mg^{2+} and radiate outward in all directions, dissipating at the outer portions of the molecule where species-specific regulatory interactions are likely to occur.

Substrate and Inhibitor Binding

The RNAP contains binding sites for two NTP substrates: the i site, which will ultimately become the 5′-end of the RNA transcript, and the i + 1 site (sometimes called the elongation site), which will extend the i site nucleotide in the 3′-direction when phosphodiester bond formation takes place. Cross-linking experiments with initiating nucleotide analogs have identified three residues that are within angstroms of the initiating nucleotide occupying the i site (Mustaev et al., 1991; Zaychikov et al., 1996) (Figure 1), βLys838 (within β_H), and βHis999 and βLys1004 (within β_I), corresponding to *E. coli* βLys1065, βHis1237, and βLys1242. These three residues are clustered together on the back wall of the RNAP channel, all no more than 11 Å from the active center Mg^{2+} (Figure 5, bottom panel). Interestingly, these residues are very highly conserved (βLys838 is absolutely conserved) but cannot play a role in the RNAP catalytic activity, since RNAP with cross-linked nucleotide adducts at these positions remains active for phosphodiester bond formation.

All of the amino acid substitutions that confer rifampicin resistance (Rif^r) to *E. coli* RNAP have been mapped to different regions of the β subunit (Jin and Gross, 1988; Severinov et al., 1993) (Figure 1). These sites cluster around a pocket on the upper wall of the main channel (Figure 5, bottom panel and Figure 7). The center of the pocket is roughly 20 Å (in a straight line) from the catalytic center Mg^{2+}, explaining how initiating nucleotide analogs covalently attached to rifampicin by a 15 Å linker arm can be active for phosphodiester bond formation (Mustaev et al., 1994). A number of observations indicate the Rif site lies along the 5′-direction upstream from the active center, near the −2 to −3 position of the DNA template strand (Mustaev et al., 1994). Consistent with this, a fourth site of cross-linking to the γ-phosphate of an initiating substrate analog (but not the α- or β-phosphate) has been mapped to a peptide fragment contained within one of the Rif regions (Severinov et al., 1995) (Figure 7a). In the presence of rifampicin, RNAP forms the open complex on promoter DNA and initiates RNA synthesis, but elongation of the RNA product halts after only a few nucleotides. Elongating RNAP, however, is resistant to rifampicin. These properties have led to the idea that the presence of rifampicin inhibits RNA synthesis by blocking the path of the elongating RNA.

DNA and RNA Interactions

To map the structural components of the RNAP involved in the formation of the template DNA and product RNA-binding sites, a series of stalled elongation complexes were analyzed in which cross-linkable probes were incorporated into specific positions of the DNA or RNA (Markovtsov et al., 1996; Nudler et al., 1996, 1998). The results of these studies are summarized in a recent review (Nudler, 1999) and mapped onto the structure in Figure 7a. In these views, the RNAP molecule has been sliced in half down the middle of the channel, then the two halves have been splayed apart (like opening a book) to view the inner surfaces of the top and bottom walls of the channel. In these views, some of the structural features discussed earlier become very apparent and are labeled. These include the wall formed by β′ regions F and G that bifurcates the main channel (causing the formation of the secondary channel), and the "rudder" formed by β'_C, which extends up from the bottom surface of the channel. On the left side of Figure 7a, the sequence conservation in β and β′ is mapped onto the exposed surfaces. On the right, the cross-link mapping studies and other information are displayed. On the right, surrounding the active center Mg^{2+} (magenta sphere) is the absolutely conserved -NADFDGD- motif, shown in red. Cross-links mapped to the α-phosphate of the initiating nucleotide occupying the i site are shown in yellow (just visible near the active center Mg^{2+} on the bottom view). Further in the 5′ direction, the γ-phosphate of the initiating nucleotide cross-links to a peptide fragment on the upper face of the channel (shown in yellow in the top view) that is coincident with the Rif site (shown in magenta in the top view). Alternatively, cross-links from the 3′-end of the RNA transcript are mapped in orange.

Crosslinks from the downstream portion of the DNA template strand (from +3 to +15, with −1 denoting the 3′-end of the RNA transcript) map to the upper surface of the channel (shown in green in the top view), while cross-links from probes incorporated further upstream on the template strand (from +12 to −4) map mainly to the bottom surface of the channel (shown in green in the bottom view). Finally, cross-links from probes incorporated at the −10 position of the RNA transcript map to a region on the bottom surface of the channel near the "rudder" (shown in blue in the bottom view).

Discussion

The features and dimensions of the core RNAP structure, along with the mapping results localizing relative positions of the nucleotide framework within the elongating RNAP, suggest a model for the transcription complex schematically illustrated in Figure 7a. All of the results mapped onto the structure in Figure 7a, including the relative orientations of 3′ and 5′ sites of the RNA transcript, as well as downstream and upstream positions of the DNA template, orient the RNAP with respect to the nucleotide framework as denoted by the arrows in Figure 7, indicating the downstream DNA entering the RNAP and the upstream DNA exiting. These considerations place the wall that bifurcates the main channel in a downstream position, and the rudder and flexible flap upstream.

The 3′-proximal 8 to 9 nucleotides of the RNA transcript (positions −1 to −9) form an RNA-DNA hybrid with the DNA template strand (Nudler et al., 1997). The cross-link from the −10 position of the RNA to a site near the upstream rudder places this protein feature near the upstream edge of the transcription bubble, where the DNA template strand is separated from the RNA transcript and reanneals with the DNA nontemplate strand. This indirectly suggests that the upstream rudder may play a role in these processes. It is interesting to note that a β hairpin loop in T7 RNAP, which is reminiscent of the upstream rudder, plays a direct role in forming the upstream edge of the transcription bubble in that system (Cheetham et al., 1999).

The RNA transcript is protected against RNAse digestion from the 3′-end to around position −14 (Komissarova and Kashlev, 1998), presumably by being bound within the ternary elongation complex. In our model, the 3′-proximal 8 to 9 nucleotides are enclosed in the main channel of the RNAP and engaged in a hybrid with the DNA template strand. This leaves single-stranded RNA from about position −10 to −14 to be accomodated within the structure. Domain 6 of the β subunit (the "flexible flap") is interesting in this regard. As discussed earlier, it forms an independent structural domain that appears to be flexibly connected to the RNAP. A groove or channel is thus formed between the domain and the RNAP (see Figure 7a), and the disposition of this groove with respect to the rudder is such that the upstream RNA could easily enter this groove and extend through it. We thus suggest that the upstream RNA (from about −10 to −14) is enclosed between the flexible flap and the main body of the RNAP, explaining the RNAse footprinting results. The possible flexibility of the flap, which would allow it to open and close (as suggested in Figure 7b), raises interesting possibilities for potential structural changes of the RNAP coincident with the transition from a promoter-specific open complex (in the presence of the σ subunit), to the highly stable and processive ternary elongation complex, and finally to a termination complex in which the nucleic acids are rapidly released.

In the downstream direction, the two DNA strands are annealed only one or two bases downstream of the RNA 3′-end. About 9 bp of the downstream double-stranded DNA are bound in the RNAP and required for the stability of the elongation complex (Nudler et al., 1996). Moreover, both strands of the DNA in this downstream region are completely protected from hydroxyl-radical cleavage, suggesting enclosure in a protein tunnel (Metzger et al., 1989; Schickor et al., 1990; Mecsas et al., 1991; Polyakov et al., 1995). Assuming the DNA is B form, about 30 Å of double-stranded DNA need to be accomodated in the channel. The length of the RNAP channel from the proposed entry point of the downstream DNA to the bifurcation point at the β'_F helix is approximately 30 Å and could thus account for these data. The DNA in this region is enclosed between the top and bottom walls of the channel, explaining the hydroxyl-radical footprinting results.

In the model, the upstream and downstream double-stranded DNA form an angle greater than 90° with respect to each other (Figure 7), closely matching the disposition of the DNA with respect to the RNAP molecule observed by atomic force microscopy of ternary elongation complexes (Rees et al., 1993). The main chamber of the RNAP channel is occupied by 9 bp of double-stranded DNA, about 9 bp of the RNA-DNA hybrid, and the nontemplate strand of the DNA, which is in some unknown location. This appears to leave little room for entry of the NTP substrates into the active center. We thus propose that the secondary channel allows access of the NTP substrates to the active center of the RNAP.

It should be noted that this model depicts the structure of the ternary elongation complex, which is able to maintain the structure of the transcription bubble and RNA-DNA hybrid as it processively translocates along the DNA template in the downstream direction. However, the core RNAP is unable to initiate the formation of this structure from a double-stranded DNA template. This function requires additional protein factors, either the promoter specificity σ subunit of prokaryotes (Helmann and Chamberlin, 1988) or a set of basal transcription factors for the eukaryotic enzymes (Conaway and Conaway, 1990). These factors are known to interact closely with both the core RNAP as well as the DNA, indicating that the structure of an initiation complex at a promoter may be very different from the model for an elongation complex presented here.

This model suggested by the structure is consistent with the available evidence discussed above. Nevertheless, it is not proven by any of the experimental data, and other models consistent with the data could be constructed. In the end, a detailed understanding of the complete transcription process will require structures of complexes between RNAP, DNA, RNA, and regulatory factors trapped at each stage of the transcription cycle for both prokaryotes and eukaryotes (Gnatt et al., 1997). We hope that the availability of the core RNAP structure and the model we have proposed will stimulate and guide further experimentation directed toward understanding transcription and its regulation.

Experimental Procedures

Purification and Crystallization

T. aquaticus frozen cell paste was a generous gift of M. Capp and T. Record (University of Wisconsin). The preperative procedure for *T. aquaticus* core RNAP was similar to the preparation of *E. coli* core RNAP (Polyakov et al., 1995). Briefly, approximately 200 g wet cell paste was thawed and lysed using a continuous-flow French press. After a low-speed spin, the soluble fraction was precipitated with 0.6% Polymin-P. RNAP was eluted from the Polymin-P pellet with TGED buffer (10 mM Tris-HCl [pH 8], 5% glycerol, 1 mM EDTA, 1 mM DTT) + 1 M NaCl, then precipitated by adding 33% (w/v) ammonium sulfate. The pellet was resuspended and loaded onto a 50 ml column of heparin-Sepharose FF (Pharmacia) equilibriated with TGED + 0.2 M NaCl. The RNAP was eluted from the column with TGED + 0.6 M NaCl. The RNAP was again precipitated with ammonium sulfate, then resuspended and loaded on a Superdex-200 gel filtration column equilibriated with TGED + 0.5 M NaCl. Fractions containing RNAP were pooled and loaded onto a Mono-Q (Pharmacia) ion-exchange column equilibriated with TGED + 0.1 M NaCl. The protein was eluted with a gradient from 0.1 to 0.5 M NaCl. The RNAP peak eluted at around 0.3 M NaCl. The RNAP was concentrated using a centrifugal filter, then loaded onto an SP Sepharose (Pharmacia) column equilibriated in TGED + 0.1 M NaCl. After loading, the column was incubated at 4°C for at least 10 hr, then pure RNAP was eluted with a 0.1 to 0.5 M NaCl gradient (core RNAP eluted at around 0.3 M NaCl). Two hundred grams of wet cell paste typically yielded 15 mg of core RNAP, which was more than

99% pure as judged from overloaded, Coomassie-stained SDS gels. This sample was ready for crystallization.

Crystals of *T. aquaticus* core RNAP were grown by vapor diffusion. Ten microliters of protein solution (17 mg/ml) were mixed with the same volume of a solution containing 40%–45% saturated $(NH_4)_2SO_4$, 0.1 M Tris-HCl (pH 8.0), and 20 mM $MgCl_2$, and incubated as a hanging drop over the same solution. Crystals grew in 2–3 weeks to typical dimensions of 0.15 mm × 0.15 mm × 0.4 mm at room temperature. For cryocrystallography, the crystals were presoaked in stabilization solution (same as the crystallization solution except with 50% saturated ammonium sulfate). The crystals were then soaked in stabilization solution containing 50% (w/v) sucrose for about 30 min before flash freezing. The frozen crystals diffracted to 4.0 Å from an in-house X-ray generator. Spots could sometimes be observed, in one direction, to 2.7 Å resolution at synchrotron beamlines. Diffraction data were processed using DENZO and SCALEPACK (Otwinowski, 1991).

Selenomethionyl core RNAP was prepared and crystallized using the same procedures from *T. aquaticus* cells grown in minimal media (culture medium 162) (Degryse et al., 1978). Cells were induced to incorporate selenomethionine by suppression of methionine biosynthesis (Doublie, 1997).

Cloning and Sequencing

Several internal fragments of the *rpoBC* operon and a fragment of *rpoA* were PCR amplified from *T. aquaticus* genomic DNA using degenerate primers designed to anneal to evolutionarily conserved segments, then cloned and sequenced. The remaining internal portions of *rpoBC* were amplified and sequenced using new primers based on the sequences obtained in the first round. The 5′- and 3′-terminal portions of the *rpo* genes were amplified using either the LA PCR in vitro cloning kit (PanVera) or by creating a genomic library and selecting clones containing relevant portions of the *rpo* genes. The DNA sequence was then determined on both strands using the fluorescent dye deoxytermination method at the Rockefeller University Protein/DNA Technology Center, or manually using the Sequenase 2.0 kit (Amersham).

Structure Determination

Patterson maps were calculated using PHASES (Furey and Swaminathan, 1997) for the ethyl-$HgCl_2$ and Ta_6Br_{14} derivatives and using the Pb derivative as native. Strong peaks (6 to 8σ) were observed on Harker sections for both derivatives at 6 Å resolution. The location of a single binding site was derived manually and confirmed using HEAVY (Terwilliger et al., 1987) for each derivative and cross-confirmed using difference Fourier methods. Additional sites, as well as sites for all the other heavy metal derivatives, were obtained using difference Fourier methods. The final phasing calculations were performed using SHARP (de La Fortelle et al., 1997). Due to errors between groups of data from each synchrotron beamline, the four data sets from CHESS A1 (Table 1) were initially refined with SHARP. Other groups of data were subsequently included but with the refined heavy-atom parameters for the previously refined data sets fixed for all subsequent refinements. After each trial refinement, density modification and phase extension from 4.5 to 3.2 Å resolution was performed using SOLOMON. Data sets were discarded and the previous refinement was used unless the new maps were noticeably improved by visual inspection. Of 40 total derivative data sets that were collected, the 9 listed in Table 1 were used for the final phase calculations.

Map interpretation and model building were done using O (Jones et al., 1991). Model building started with the α subunits, the fold of which was immediately recognized from the previously solved *E. coli* αNTD (Zhang and Darst, 1998). Preliminary rounds of α refinement were performed by creating a solvent mask around the α model, cutting out the electron density map inside the volume of the α mask, then back-transforming the resulting electron density map. The resulting structure factors were used for two rounds of refinement of the α structure. Subsequently, initial refinements of the entire RNAP model were performed by keeping the α coordinates fixed. Only in the last round of positional refinement was α refined along with the rest of the RNAP model (but with tight noncrystallographic restraints between the appropriate α domains). Refinement calculations were performed using CNS (Adams et al., 1997). From an initial R factor of 0.44 (R_{free} = 0.45), the current R factor is 0.35 (R_{free} = 0.41) for data from 100–3.2 Å resolution and a 0σ cutoff (with bulk solvent correction and group β factor refinement), 0.33 for data from 8–3.3 (R_{free} = 0.39). The R_{free} was closely monitored during all refinement procedures. Further refinement is in progress.

Acknowledgments

We thank M. Capp and T. Record for the gift of frozen *T. aquaticus* cell paste; A. Mustaev for stimulating discussions regarding the structure and for experiments to confirm the binding of Pb in the active center metal site; and V. Nikiforov for the extensive alignment of RNAP subunits and for discussions regarding the structure. Some of the DNA and protein sequencing was performed by the Rockefeller University Protein/DNA Technology Center. We are indebted to J. Bonanno and D. Jeruzalmi for assistance with crystallographic calculations; R. Bennet of the Rockefeller University Computing Service for invaluable assistance; J. Kappler and P. Marrack for providing computing facilities; D. Thiel, S. Gruner, and members of the MacCHESS staff, L. Berman and the staff at NSLS X25, and W. Schildkamp and the BioCARS staff at APS for support and assistance in data collection; G. Schneider for the gift of Ta_6Br_{14}; W. Jahn for the Ir_4 cluster; and J. Goldberg, R. Landick, and S. Nair for helpful discussions. We are indebted to S. K. Burley, J. Kuriyan, R. MacKinnon, and the rest of the Rockefeller University structural biology community for making this project possible. S. A. D. is especially grateful to R. Kornberg, whose own work is a constant inspiration. G. Z. was supported by a National Research Service Award (NIH GM19441-01). E. C. was supported by a Kluge Postdoctoral Fellowship. This work was supported in part by a Burroughs Wellcome Career Development Award to K. S. and a Pew Scholars Award in the Biomedical Sciences to S. A. D.

Received September 2, 1999; revised September 25, 1999.

References

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.

Archambault, J., and Friesen, J.D. (1993). Genetics of RNA polymerases I, II, and III. Microbiol. Rev. *57*, 703–724.

Borukhov, S., Severinov, K., Kashlev, M., Lebedev, A., Bass, I., Rowland, G.C., Lim, P.-P., Glass, R.E., Nikiforov, V., and Goldfarb, A. (1991). Mapping of trypsin cleavage and antibody-binding sites and delineation of a dispensable domain in the β subunit of *Escherichia coli* RNA polymerase. J. Biol. Chem. *266*, 23921–23926.

Chamberlin, M., and Berg, B. (1962). Deoxyribonucleic acid-directed synthesis of ribonucleic acid by an enzyme from *Escherichia coli*. Proc. Natl. Acad. Sci. USA *48*, 81–94.

Cheetham, G.M.T., Jeruzalmi, D., and Steitz, T.A. (1999). Structural basis for initiation of transcription from an RNA polymerase-promoter complex. Nature *399*, 80–83.

Conaway, J.W., and Conaway, R.C. (1990). An RNA polymerase II transcription factor shares functional properties with *Escherichia coli* sigma 70. Science *248*, 1550–1553.

Darst, S.A., Kubalek, E.W., and Kornberg, R.D. (1989). Three-dimensional structure of Escherichia coli RNA polymerase holoenzyme determined by electron crystallography. Nature *340*, 730–732.

Darst, S.A., Edwards, A.M., Kubalek, E.W., and Kornberg, R.D. (1991). Three-dimensional structure of yeast RNA polymerase II at 16 Å resolution. Cell *66*, 121–128.

Darst, S.A., Polyakov, A., Richter, C., and Zhang, G. (1998a). Insights into Escherichia coli RNA polymerase structure from a combination of X-ray and electron crystallography. J. Struct. Biol. *124*, 115–122.

Darst, S.A., Polyakov, A., Richter, C., and Zhang, G. (1998b). Structural studies of *Escherichia coli* RNA polymerase. Cold Spring Harb. Symp. Quant. Biol. *63*, 269–276.

Degryse, E., Glansdorff, N., and Pierard, A. (1978). A comparative

analysis of extreme thermophilic bacteria belonging to the genus *Thermus*. Arch. Microbiol. *117*, 189–196.

de La Fortelle, E., Irwin, J.J., and Bricogne, G. (1997). SHARP: a maximum-likelihood heavy-atom parameter refinement and phasing program for the MIR and MAD methods. In Crystallographic Computing, P. Bourne and K. Watenpaugh, eds., pp. 1–9.

Doublie, S. (1997). Preparation of selenomethionyl proteins for phase determination. Methods Enzymol. *276*, 523–530.

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.

Furey, W., and Swaminathan, S. (1997). PHASES-95: a program package for the processing and analysis of diffraction data from macromolecules. Methods Enzymol. *277*, 590–620.

Gentry, D.R., and Burgess, R.R. (1993). Cross-linking of *Escherichia coli* RNA polymerase subunits: identification of β′ as the binding site of σ. Biochemistry *32*, 11224–11227.

Gnatt, A., Fu, J., and Kornberg, R.D. (1997). Formation and crystallization of yeast RNA polymerase II elongation complexes. J. Biol. Chem. *272*, 30799–30805.

Gross, C.A., Chan, C.L., and Lonetto, M.A. (1996). A structure/function analysis of *Escherichia coli* RNA polymerase. Phil. Trans. Royal Soc. Lond. B: Biol. Sci. *351*, 475–482.

Helmann, J.D., and Chamberlin, M.J. (1988). Structure and function of bacterial sigma factors. Annu. Rev. Biochem. *57*, 839–872.

Heyduk, T., Heyduk, E., Severinov, K., Tang, H., and Ebright, R.H. (1996). Rapid epitope mapping by hydroxyl-radical protein footprinting: determinants of RNA polymerase alpha subunit for interaction with β, β′, and σ subunits. Proc. Natl. Acad. Sci. USA *93*, 10162–10166.

Huang, R.C., Maheshwari, N., and Bonner, J. (1960). Enzymatic synthesis of RNA. Biochem. Biophys. Res. Commun. *3*, 689.

Hurwitz, J., Bresler, A., and Diringer, R. (1960). The enzymatic incorporation of ribonucleotides into polyribonucleotides and the effect of DNA. Biochem. Biophys. Res. Commun. *3*, 15.

Jin, D.J., and Gross, C.A. (1988). Mapping and sequencing of mutations in the *Escherichia coli rpoB* gene that lead to rifampicin resistance. J. Mol. Biol. *202*, 45–58.

Jokerst, R.S., Weeks, J.R., Zehring, W.A., and Greenleaf, A.L. (1989). Analysis of the gene encoding the largest subunit of RNA polymerase II in *Drosophila*. Mol. Gen. Genet. *215*, 266–275.

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

Komissarova, N., and Kashlev, M. (1998). Functional topography of nascent RNA in elongation intermediates of RNA polymerase. Proc. Natl. Acad. Sci. USA *95*, 14699–14704.

Lee, B., and Richards, F.M. (1971). The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. *55*, 379–400.

Markov, D., Naryshkina, T., Mustaev, A., and Severinov, K. (1999). A zinc binding site in the largest subunit of DNA-dependent RNA polymerase is involved in enzyme assembly. Genes Dev., in press.

Markovtsov, V., Mustaev, A., and Goldfarb, A. (1996). Protein-RNA interactions in the active center of transcription elongation complex. Proc. Natl. Acad. Sci. USA *93*, 3221–3226.

Mecsas, J., Cowing, D.W., and Gross, C.A. (1991). Development of RNA polymerase-promoter contacts during open complex formation. J. Mol. Biol. *220*, 585–597.

Metzger, W., Schickor, P., and Heumann, H. (1989). A cinematographic view of *Escherichia coli* RNA polymerase translocation. EMBO J. *8*, 2745–2754.

Mukherjee, K., and Chatterji, D. (1997). Studies on the omega subunit of *Escherichia coli* RNA polymerase: its role in the recovery of denatured enzyme activity. Eur. J. Biochem. *247*, 884–889.

Mustaev, A., Kashlev, M., Lee, J., Polyakov, A., Lebedev, A., Zalenskaya, K., Grachev, M., Goldfarb, A., and Nikiforov, V. (1991). Mapping of the priming substrate contacts in the active center of *Escherichia coli* RNA polymerase. J. Biol. Chem. *266*, 23927–23931.

Mustaev, A., Zaychikov, E., Severinov, K., Kashlev, M., Polyakov, A., Nikiforov, V., and Goldfarb, A. (1994). Topology of the RNA polymerase active center probed by chimeric rifampicin-nucleotide compounds. Proc. Natl. Acad. Sci. USA *91*, 12036–12040.

Mustaev, A., Kozlov, M., Markovtsov, V., Zaychikov, E., Denissova, L., and Goldfarb, A. (1997). Modular organization of the catalytic center of RNA polymerase. Proc. Natl. Acad. Sci. USA *94*, 6641–6645.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins: Struct. Funct. Genet. *11*, 281–296.

Nudler, E. (1999). Transcription elongation: structural basis and mechanisms. J. Mol. Biol. *288*, 1–12.

Nudler, E., Avetissova, E., Markovtsov, V., and Goldfarb, A. (1996). Transcription processivity: protein-DNA interactions holding together the elongation complex. Science *273*, 211–217.

Nudler, E., Mustaev, A., Lukhtanov, E., and Goldfarb, A. (1997). The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. Cell *89*, 33–41.

Nudler, E., Gusarov, I., Avetissova, E., Kozlov, M., and Goldfarb, A. (1998). Spatial organisation of transcription elongation complex in *Escherichia coli*. Science *281*, 424–428.

Otwinowski, Z. (1991). Isomorphous Replacement and Anomalous Scattering. W. Wolf, P.R. Evans, and A.G.W. Leslie, eds. (Daresbury, UK: Science and Engineering Research Council, Daresbury Laboratory).

Polyakov, A., Severinova, E., and Darst, S.A. (1995). Three-dimensional structure of *Escherichia coli* core RNA polymerase: promoter binding and elongation conformations of the enzyme. Cell *83*, 365–373.

Rees, W.A., Keller, R.W., Vesenka, J.P., Yang, G., and Bustamante, C. (1993). Evidence of DNA bending in transcription complexes imaged by scanning force microscopy. Science *260*, 1646–1649.

Rost, B., and Sander, C. (1993). Prediction of protein structure at better than 70% accuracy. J. Mol. Biol. *232*, 584–599.

Schickor, P., Metzger, W., Wladyslaw, W., Lederer, H., and Heumann, H. (1990). Topography of intermediates in transcription initiation of *E. coli*. EMBO J. *9*, 2215–2220.

Schultz, P., Celia, H., Riva, M., Sentenac, A., and Oudet, P. (1993). Three-dimensional model of yeast RNA polymerase I determined by electron microscopy of two-dimensional crystals. EMBO J. *12*, 2601–2607.

Sentenac, A., Riva, M., Thuriaux, P., Buhler, J.-M., Treich, I., Carles, C., Werner, M., Ruet, A., Huet, J., Mann, C., et al. (1992). Yeast RNA polymerase subunits and genes. In Transcriptional Regulation, S.L. McKnight and K.R. Yamamoto, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory), pp. 27–54.

Severinov, K., Mustaev, A., Kashlev, M., Borukhov, S., Nikiforov, V., and Goldfarb, A. (1992). Dissection of the β subunit in the *Escherichia coli* RNA polymerase into domains by proteolytic cleavage. J. Biol. Chem. *267*, 12813–12819.

Severinov, K., Soushko, M., Goldfarb, A., and Nikiforov, A. (1993). Rifampicin region revisited. New rifampicin-resistant and streptolydigin-resistant mutants in the beta subunit of *Escherichia coli* RNA polymerase. J. Biol. Chem. *268*, 14820–14825.

Severinov, K., Soushko, M., Goldfarb, A., and Nikiforov, V. (1994). RifR mutations in the beginning of the *Escherichia coli rpoB* gene. Mol. Gen. Genet. *244*, 120–126.

Severinov, K., Mustaev, A., Severinova, E., Kozlov, M., and Darst, S.A., and Goldfarb, A. (1995). The β subunit Rif-cluster I is only angstroms away from the active center of *Escherichia coli* RNA polymerase. J. Biol. Chem. *270*, 29428–29432.

Severinov, K., Mooney, R., Darst, S.A., and Landick, R. (1997). Tethering of the large subunits of *Escherichia coli* RNA polymerase. J. Biol. Chem. *272*, 24137–24140.

Stevens, A. (1960). Incorporation of the adenine ribonucleotide into RNA by cell fractions from *E. coli* B. Biochem. Biophys. Res. Commun. *3*, 92.

Sweetser, D., Nonet, M., and Young, R.A. (1987). Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. Proc. Natl. Acad. Sci. USA *84*, 1192–1196.

Terwilliger, T.C., Kim, S.H., and Eisenberg, D.E. (1987). Generalized method of determining heavy atom positions from the difference Patterson function. Acta Crystallogr. A *43*, 34–38.

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*, 389–446.

Wang, Y., Severinov, K., Loizos, N., Fenyö, D., Heyduk, E., Heyduk, T., Chait, B.T., and Darst, S.A. (1997). Determinants for *Escherichia coli* RNA polymerase assembly within the β subunit. J. Mol. Biol. *270*, 648–662.

Weiss, S., and Gladstone, L. (1959). A mammalian system for the incorporation of cytidine and triphosphate into ribonucleic acid. J. Am. Chem. Soc. *81*, 4118–4119.

Zakharova, N., Paster, B.J., Wesley, I., Dewhirst, F.E., Berg, D.E. and Severinov, K. (1999). Fused and overlapping rpoB and rpoC genes in helicobacters, campylobacters, and related bacteria. J. Bacteriol. *181*, 3857–3859.

Zaychikov, E., Martin, E., Denissova, L., Kozlov, M., Markovtsov, V., Kashlev, M., Heumann, H., Nikiforov, V., Goldfarb, A., and Mustaev, A. (1996). Mapping of catalytic residues in the RNA polymerase active center. Science *273*, 107–109.

Zhang, G., and Darst, S.A. (1998). Structure of the *Escherichia coli* RNA polymerase a subunit amino-terminal domain. Science *281*, 262–266.

Protein Data Bank ID Code

The coordinates for the sequence described in this work have been deposited and are available from the author.