Cell, Vol. 119, 481–489, November 12, 2004, Copyright ©2004 by Cell Press

Structural Basis of Transcription: Nucleotide Selection by Rotation in the RNA Polymerase II Active Center

Kenneth D. Westover, David A. Bushnell, and Roger D. Kornberg*
Department of Structural Biology
Stanford University School of Medicine
Stanford, California 94305

Summary

Binding of a ribonucleoside triphosphate to an RNA polymerase II transcribing complex, with base pairing to the template DNA, was revealed by X-ray crystallography. Binding of a mismatched nucleoside triphosphate was also detected, but in an adjacent site, inverted with respect to the correctly paired nucleotide. The results are consistent with a two-step mechanism of nucleotide selection, with initial binding to an entry (E) site beneath the active center in an inverted orientation, followed by rotation into the nucleotide addition (A) site for pairing with the template DNA. This mechanism is unrelated to that of single subunit RNA polymerases and so defines a new paradigm for the large, multisubunit enzymes. Additional findings from these studies include a third nucleotide binding site that may define the length of backtracked RNA; DNA double helix unwinding in advance of the polymerase active center; and extension of the diffraction limit of RNA polymerase II crystals to 2.3 Å.

Introduction

The elementary step in transcription may be subdivided into multiple stages: selection of a ribonucleoside triphosphate complementary to the DNA template; catalysis of phosphodiester bond formation; and translocation of the RNA and DNA, with concomitant unwinding of the RNA-DNA hybrid helix and unwinding and rewinding of the DNA double helix. Previous X-ray crystal structures of yeast RNA polymerase II (pol II) transcribing complexes have given insight into the mechanisms of translocation and helix unwinding (Gnatt et al., 2001; Westover et al., 2004). We now report structures of transcribing complexes that are informative about the mechanisms of nucleotide selection and catalysis.

In the first structure of a pol II transcribing complex (Gnatt et al., 2001), transcription was stalled by the omission of a nucleoside triphosphate (NTP). The structure revealed the last nucleotide added to the RNA, still in the "nucleotide addition" (here designated "A") site, opposite the DNA base at position i+1 in the template strand. The structure therefore represented the "pretranslocation" state (although the complex must have undergone translocation, advancing the last nucleotide added and the associated DNA base to position i+2, exposing the A site to create a requirement for the missing NTP, and stalling transcription; the complex must then have undergone backtracking to return the last nucleotide to the A site). In the structure of this transcribing complex, the DNA base at i+1 is contacted by amino acid side chains of a structural element termed the "bridge helix." In pol II, the bridge helix is essentially straight, whereas in the structure of bacterial RNA polymerase (Zhang et al., 1999), the bridge helix is bent, placing the corresponding amino acid side chains in position to contact the DNA base at position i+2. This observation led to the proposal that transitions between straight and bent states of the bridge helix underlie the translocation step in transcription (Cramer et al., 2001). Evidence in support of this hypothesis has come from RNA-protein crosslinking studies, in which crosslinks characteristic of both straight and bent states of the bacterial enzyme were observed (Epshtein et al., 2002).

A second structure of a pol II transcribing complex was obtained with the use of RNA formed not by transcription, but rather as a synthetic oligonucleotide (Westover et al., 2004). An eight- or nine-residue RNA and complementary strand of DNA form a stable complex with pol II (Kireeva et al., 2000). A chain terminating 3′-deoxyadenylate residue was added by transcription. The structure of this transcribing complex revealed a vacant A site and therefore represented the "posttranslocation" state. The structure further revealed the unwinding of the DNA-RNA hybrid at the upstream end of the RNA, involving a set of protein loops in a network of protein-nucleic acid interactions.

We have now exploited the formation of a transcribing complex in the posttranslocation state to investigate the mode of interaction with substrate NTP. Transcribing complex crystals were soaked with both NTP matched to the DNA base at position i+1 and unmatched NTPs. The resulting structures revealed an unexpected feature of the nucleotide entry and addition mechanism.

Results

We previously formed a transcribing complex through the binding to pol II of a nine-residue RNA oligonucleotide and a 15-residue DNA oligonucleotide containing complementary sequence (Westover et al., 2004). We have now extended the downstream region of the template strand with 14 residues of duplex DNA (Figure 1), improving the reproducibility and size of the crystals and their stability to manipulation before freezing. Introduction of a noncomplementary NTP before freezing gave the best diffraction data. Structures were solved by molecular replacement with an earlier transcribing complex model (Gnatt et al., 2001) and rigid body refinement (Table 1).

DNA Unwinding in the Transcribing Complex

As with the previous transcribing complex formed from RNA and DNA oligonucleotides, the present transcribing complex was in the posttranslocation state. The structure differed from those obtained previously by the presence of connected density for the template strand beyond the downstream end of the DNA-RNA hybrid

*Correspondence: kornberg@stanford.edu

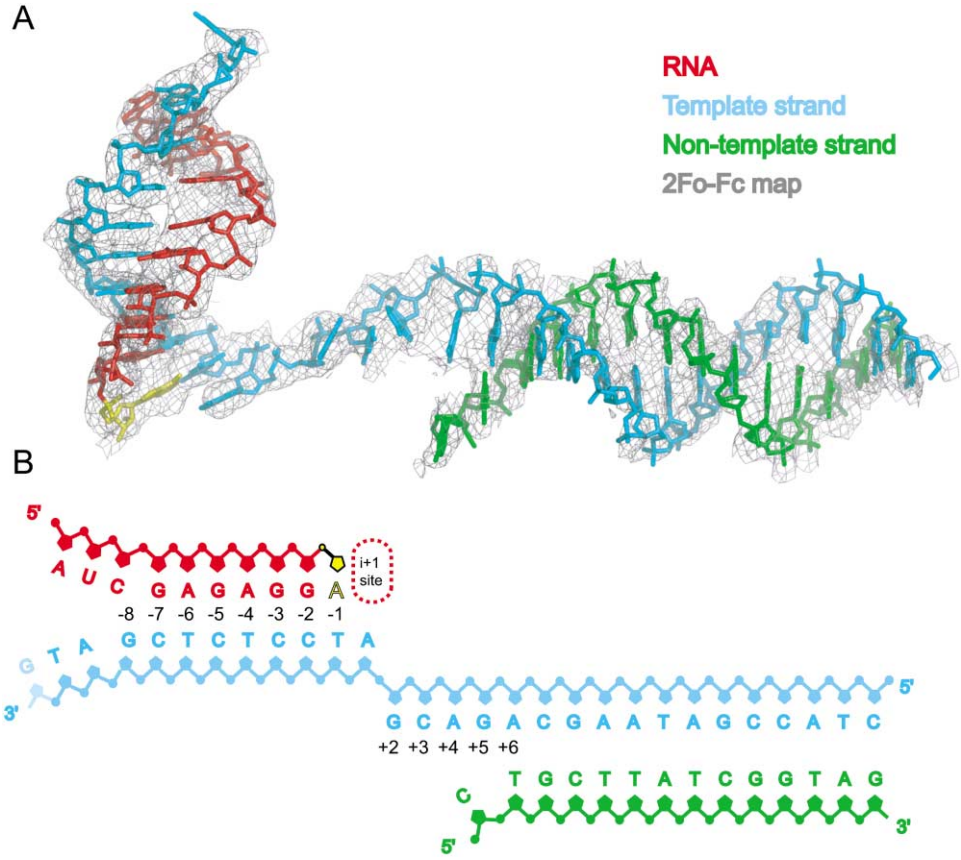

Figure 1. DNA and RNA in the Structure of a Pol II Transcribing Complex

(A) Model for DNA and RNA fitted to electron density for nucleic acids ($2F_o-F_c$ map, with phases from pol II alone, contoured at 0.8 σ). View is the same as Figure 1 of Westover et al. (2004). Color code is at upper right. A chain-terminating 3′ dA residue is shown in yellow.

(B) Sequences of DNA and RNA in the transcribing complex. Color code as in (A). Nucleotide positions are numbered with respect to the addition site at +1 (denoted i+1 site), with positions upstream extending from −1 and those downstream from +2. Separation of DNA and RNA strands upstream of −8 and separation of DNA strands upstream of +6 are shown schematically. Figures were generated by PyMOL or SPOCK.

(Figure 1). The downstream double-stranded DNA region was also better defined, lying on the Rpb1 side of the active center cleft (not shown). Beyond the template residue at i+1, additional unpaired bases were resolved on the template strand at the downstream positions +2, +3, and +4. The next base on the template strand at position +5 was also unpaired, despite the presence of a complementary base in the nontemplate strand. This "fraying" of the end of the downstream duplex was maintained by interaction of the nontemplate residue at position +5 with Rpb1 residues Lys1109 and Asn1110 in the floor of the active center cleft. Fraying of the end may have been due, in part, to the absence of any nontemplate residues beyond position +5, and indeed biochemical evidence points to strand separation further upstream during transcription of fully double-stranded DNA (Santangelo and Roberts, 2004; Shi et al., 1988). Stable fraying is, nonetheless, a significant result, and the interactions involved may contribute to the unwinding of the template DNA at the downstream end of the transcription bubble.

Binding of a Correctly Matched NTP in the A Site

Transcribing complex crystals in the posttranslocation state, with a chain-terminating residue at the 3′ end of the RNA, could be soaked in Mn- or Mg-UTP, complementary to the DNA base at position i+1, without addition of U to the RNA. Diffraction data were collected to 4.2 Å resolution (Table 1), and a difference electron density map was calculated by subtracting the structure with the DNA and RNA removed ($2F_o-F_c$ omit map). The difference map showed density for UTP in the A site (Figure 2A), with the nucleotide base paired to the adjacent template DNA base and with the α phosphate positioned for in-line nucleophilic attack by an OH group at the 3′ end of the RNA (absent from the chain-terminated RNA).

An anomalous difference map obtained from a crystal soaked in Mn-UTP showed density for two Mn ions, one coordinated by the α phosphate of UTP and by Rpb1 residues Asp481, Asp483, and Asp485 (Figure 3A). This Mn ion corresponded to "metal A" seen previously in the structure of pol II alone (Cramer et al., 2001). The second Mn ion was coordinated by the β and γ phosphates of UTP, by Rpb1 residues Asp481 and Asp483, and by Rpb2 residue Asp836. The location and coordination of this second Mn ion differed significantly from that of "metal B" in the structure of pol II alone. Whereas the previous metal B showed a low occupancy, and is of uncertain relevance to transcription, the present metal

Table 1. Crystallographic Data and Structure Statistics

	Elongation Complex					
Nucleotide soak	ATP (mismatch)	UTP (match)				
Space group	C2					
Unit cell dimensions	169 × 222 × 194 Å; 90°, 101°, 90°					
Wavelength (Å)	0.979					
Resolution (Å)	40–3.5 (3.6–3.5)	40–4.2 (4.3–4.2)				
Unique reflections	83,860	50,043				
Completeness (%)	94.2 (87.3)	95.8 (93.2)				
Redundancy	2.8	2.5				
I/Isigma	15.1	11.8				
Mosaicity (°)	0.75	0.89				
R_{sym} (%)	11.1 (37.0)	15.7 (36.1)				
Refinement						
R_{cryst}/R_{free}	23/31.7	34.9/39.8				
PDB accession	1R9T	1R9S				
	Pol II					
Nucleotide soak	UTP	GTP	ATP	CTP	2′ dATP	UTP
Space group	I222					
Unit cell dimensions	123 × 223 × 374 Å; 90°, 90°, 90°					
Wavelength (Å)	0.979					1.283
Resolution (Å)	40–2.3 (2.38–2.3)	40–3.0 (3.1–3.0)	40–3.2 (3.3–3.2)	40–3.3 (3.4–3.3)	40–3.4 (3.5–3.4)	40–3.5 (3.6–3.5)
Unique reflections	211,633	99,762	85,066	63,673	70,366	64,710
Completeness (%)	92.8 (85.5)	96.9 (85.0)	100 (100)	97.4 (98.2)	99.8 (99.8)	100 (100)
Redundancy	3.5	3.5	3.9	2.4	3.9	6.9
I/Isigma	15.3	9.4	7.3	5.7	7.0	9.2
Mosaicity (°)	0.58	0.35	0.35	0.33	0.28	0.60
R_{sym} (%)	9.8 (38.5)	9.0 (35.7)	9.5 (33.7)	13.2 (33.2)	10.4 (30.2)	10.6 (27.5)
Refinement						
R_{cryst}/R_{free}	24.7/29.4	23.4/25.5	22.7/24.6	28.1/26.3	22.2/26.2	NA
PDB accession	1TWF	1TWC	1TWA	1TWG	1TWH	NA

B was comparable in level to metal A. The present metal B stood in a similar (though not identical—see below) spatial relationship to metal A and nucleotide as reported for the two metals and nucleotide in single subunit RNA and DNA polymerases (Doublie et al., 1998; Steitz, 1998). We conclude that a second metal ion enters the transcribing complex with the substrate NTP, enabling catalysis by a two-metal ion mechanism.

The basis of specificity for a ribose sugar of the NTP in the i+1 site was not immediately apparent. Neither the highly conserved Rpb1 residue Asn479, 5 Å from the 2′ OH group of the NTP, nor the highly conserved Arg446, 4.2 Å away, is close enough for hydrogen bonding. These residues are in similar locations to Asp812 and Arg425 of T7 RNA polymerase. Mutations of the T7 enzyme at these positions abolish enzyme activity (Bonner et al., 1994), so the roles in nucleotide sugar specificity could not be assessed. Mutation of an entirely different residue, Tyr639P, eliminates specificity, but Tyr639 does not have a convincing structural counterpart in pol II (Huang et al., 1997). Some have suggested that nearby Rpb1 Tyr836 could play the same role as T7 Tyr639. However, Rpb1 Tyr836 interacts with the template strand at positions +2 and +3 and not with

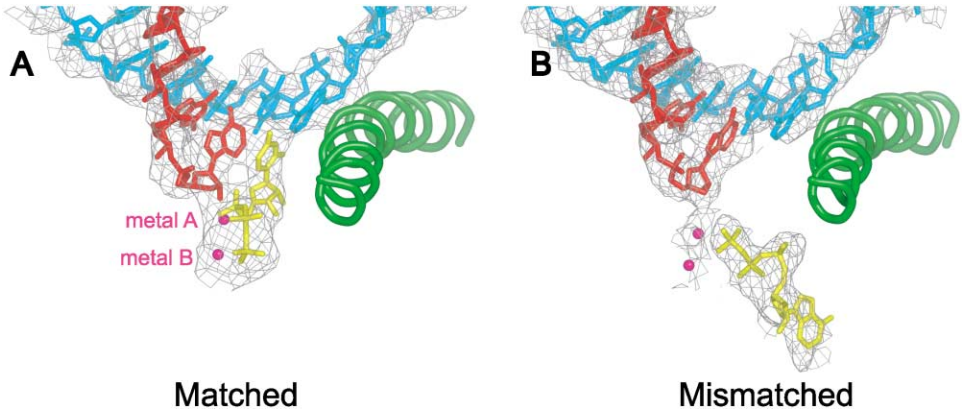

Figure 2. Downstream End of the DNA-RNA Hybrid in Transcribing Complex Structures, Showing Occupancy of the A and E Sites

(A) Transcribing complex with matched NTP (UTP) in the A site.

(B) Transcribing complex with mismatched NTP (ATP) in the E site. Views are the same as in Figure 1. DNA is blue, RNA is red, and NTPs are in yellow. Mg ions are shown as magenta spheres.

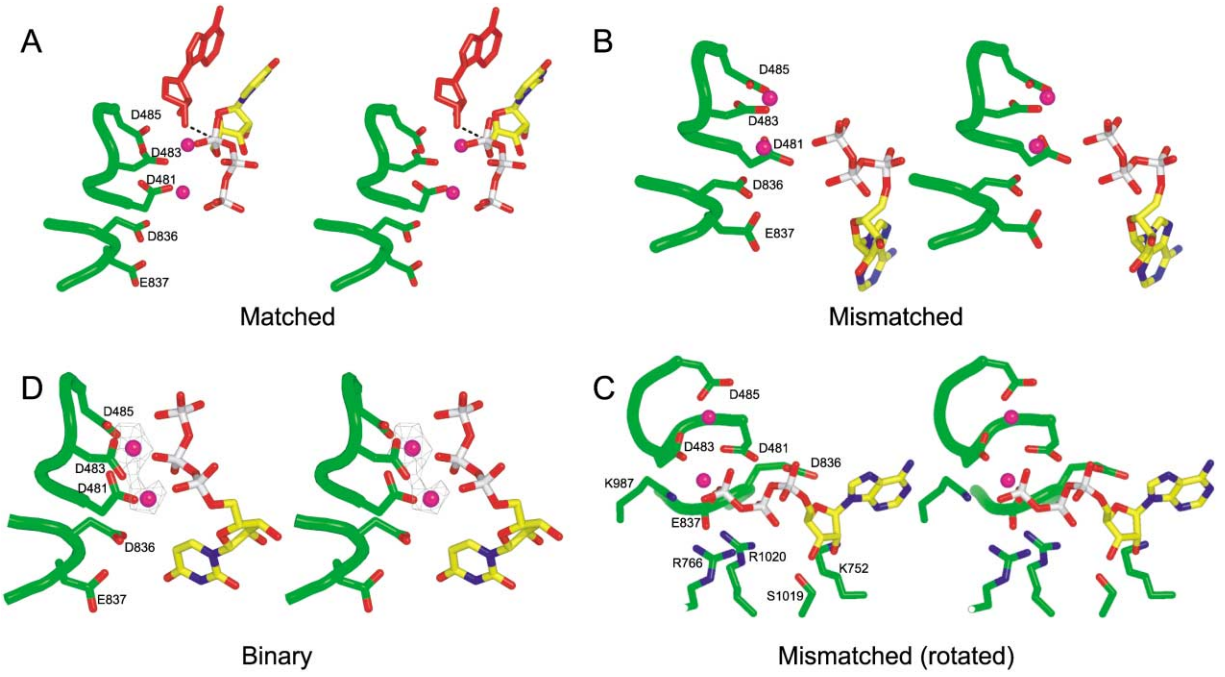

Figure 3. Two Metal Ions at the Active Center of Pol II with Bound NTPs

(A) Transcribing complex with matched NTP (UTP), same as Figure 2A.

(B and C) Transcribing complex with mismatched NTP (ATP), same as Figure 2B.

(D) Pol II alone with bound UTP. Stereo pairs are shown. The view in (A), (B), and (D) is the same as in Figure 1 and is rotated in (C) to reveal additional interacting amino acid side chains. Polypeptide chain and carbon atoms of side chains are green, oxygen atoms of side chains are red, nitrogen atoms of side chains are blue, NTPs are colored with carbons in yellow, oxygens in red, nitrogen in blue, and phosphates in white, and the 3′-terminal nucleotide of the RNA in (A) is red. Mg ions are depicted as magenta spheres. Mn anomalous difference map is shown in (D) as a gray net. Positioning of the α phosphate of the NTP in (A) for in-line nucleophilic attack by the 3′ OH group of the RNA is indicated by a dashed line.

the base in the i+1 site, as does T7 Tyr639 (Temiakov et al., 2004; Yin and Steitz, 2004).

It has also been suggested that bending of the bridge helix, as observed in bacterial RNA polymerase structures (see above), could bring Rpb1 Thr831 into position to interact with the 2′ OH of the incoming nucleotide. No such interaction or bending of the bridge helix is seen in our structure, and moreover, the bending observed in bacterial RNA polymerase would move Thr831 further from a nucleotide in the A site. In the absence of interactions at the 2′ position of the pentose ring in our structure, rNTP/dNTP discrimination may occur in the transition state, which could involve bridge helix movement in a direction different from that observed for bacterial RNA polymerase.

Binding of a Mismatched Nucleotide in the E Site

As a control, the analysis was repeated for crystals soaked with a mismatched rather than matched nucleotide. Diffraction extended to 3.5 Å and the resulting $2F_o-F_c$ omit map contained no density in the A site, showing that the correctly matched nucleotide bound specifically to that site. There was, however, density for the mismatched nucleotide in a distinct but overlapping site (Figure 2B). The orientation of the mismatched nucleotide was flipped, with β and γ phosphates coordinating metal B (Figure 3B) and the sugar and base projecting downwards into the pore beneath the active center. The phosphates and sugar interacted with Rpb1 residue Lys752 and with Rpb2 residues Arg766, Tyr769, Lys 987, Ser1019, and Arg1020. The base projected into solution, with no apparent protein contact.

As we were exploring the nature of the mismatched nucleotide site, its existence was proposed by Sosunov et al. (2003) on the basis of biochemical evidence. They showed that a mismatched nucleotide could stimulate an exonuclease activity of bacterial RNA polymerase by recruitment of a second metal ion to the active center. They deduced the location of this metal ion and proposed a role in RNA synthesis. While their specific proposals for the orientation of the bound nucleotide and its interactions with amino acid side chains are not supported by our findings, their essential ideas are correct. As discussed below, the mismatched nucleotide site facilitates the entry of NTPs to the A site, and so, in keeping with the designation by Sosunov et al., we refer to the mismatched site as the "E" site.

The positions of metals A and B vary slightly from one pol II-NTP complex to another (Figure 3), revealing some flexibility of the active site structure. This variation is not surprising—an NTP is required to stabilize metal B, so if the NTP changes position, the associated metal would be expected to change position as well. The variation in metal ion position is also consistent with the energetics of Mg^{2+} coordination. Substitution of more than three waters in the Mg^{2+} coordination sphere with

non-water ligands, in this case aspartate residues and phosphates, is disfavored, and the resulting structure will therefore be relatively unstable and prone to movement (Dudev and Lim, 2003). This flexibility is also consistent with the idea that pol II has a single tunable active site capable of operating in various modes: as a polymerase, exonuclease, endonuclease, exopyrophosphorylase, and endopyrophosphorylase (Sosunov et al., 2003).

NTP Binding to Free Pol II: Implications for Backtracking; and Extension of Diffraction from Pol II Crystals

Because mismatched NTPs interact only with protein residues and active site metals, and not with the DNA-RNA hybrid, it might be expected that NTPs would bind to free pol II. This possibility was investigated by soaking ten-subunit pol II crystals with 15 mM NTPs in the presence of 50 mM Mg or Mn chloride. The structures, solved by molecular replacement (Table 1), revealed the binding of all NTPs (GTP, CTP, ATP, UTP, and 2′dATP) to crystals of free pol II. Data from crystals soaked in Mn-UTP were of particularly high quality, extending to 2.3 Å resolution, well beyond the limit previously obtained for pol II (Cramer et al., 2001). Minor modification and refinement of the pol II model gave an R factor to this resolution similar to that previous obtained at 2.8 Å, but with the overall B factor reduced from 64 to 47 $Å^2$ (Table 1).

All NTPs bound to free pol II in the same orientation and nearly the same location as the mismatched NTP in the transcribing complex. In the absence of DNA and RNA, however, the phosphates could coordinate more extensively with metal A, so the NTP was shifted toward metal A (Figure 3C). High-quality Mn anomalous data showed the same location of metal B as in the transcribing complex, with nearly equivalent occupancies of metals A and B.

For all NTPs in all crystal soaks, there was extra density in $2F_o-F_c$ maps in the pore beneath the active center, at the interface of Rpb1 and Rpb5 (Figure 4). An NTP could be fit into this density, although the detail was insufficient for accurate modeling. The binding pocket (yellow surface in Figure 4) was formed by Rpb1 residues Tyr1349, Tyr1353, Trp954, and Phe947 and by Rpb5 residues Arg200, Lys201, Ser202, and Glu203, all of which showed especially well-defined amino acid side chain density. The aromatic residues of the pocket might interact with the RNA bases and the charged and polar residues with the phosphates. The distance from the pocket to the active center corresponds to a length of about nine residues of RNA. Thus if the 3′ end of a backtracked RNA were bound in the pocket (as modeled in Figure 4), an RNA fragment of about this length would be produced by cleavage at the active center induced by elongation factors such as SII (TFIIS) in eukaryotes and GreA/B in bacteria. Indeed cleavage products of this length are most often observed (Fish and Kane, 2002; Gu and Reines, 1995), and backtracked RNA as modeled here (Figure 4) is compatible with the structures of TFIIS and GreB in the pol II and bacterial RNA polymerase pores (Kettenberger et al., 2003; Opalka et al., 2003). The binding pocket could correspond to the previously proposed, allosteric-activating NTP binding site (Foster et al., 2001). An NTP in the pocket might prevent the stabilization of a backtracked state, leading to a larger proportion of polymerases in the actively transcribing state.

Discussion

The principal finding from this work is the direct observation of two NTP binding sites in transcribing pol II, termed A and E sites. The existence of the A site was assumed because of its necessity in the transcription mechanism. The existence of the E site was unexpected; it has also been inferred from the Mg ion dependence of an exonucleolytic activity of pol II (Sosunov et al., 2003). While the A site is found only in transcribing pol II, the E site also occurs in the free enzyme. Conversely, a Mg ion associated with the A site is a permanent component of free pol II, whereas a Mg ion associated with the E site enters with the NTP at that site.

The A and E sites overlap and so cannot be simultaneously occupied. Both sites also overlap with the acidic hairpin of TFIIS bound to pol II (Kettenberger et al., 2003). By contrast, a recently described ppGpp site in bacterial RNA polymerase does not overlap the A site (Artsimovitch et al., 2004), and moreover, ppGpp binding is accompanied by three Mg ions, in roughly the same positions as the two reported here and that reported previously (Cramer et al., 2001). (It may be noted that the three Mg ions are seen in only one of two molecules in the unit cell, and that 362 Mg ions were included in the structure by the use of a sigma-cutoff criterion for designation as Mg, in contrast with the collection of anomalous diffraction data from crystals soaked in Mn as in our work.)

Functional Significance of the E Site

How might NTP binding at the E site contribute to the transcription process? NTPs may bind at the E site and then rotate around metal B to sample base pairing in the A site. Such binding doubtless occurs because nucleotides first encounter the E site upon entry through the pore (Figure 5A). Nucleotide rotation may also play a role in rNTP/dNTP discrimination, as no selection mechanism is apparent from our structures of nucleotides stably bound in the A and E sites.

These ideas have been pursued by a theoretical analysis of NTP diffusion in the pol II pore (N. Batada et al., submitted). The results show a significant restriction of NTP diffusion by the pore. The restriction is largely overcome by an enhanced lifetime of NTPs in the pore, due in part to binding at the E site.

Comparison of Pol II with Small Single Subunit Polymerases

Pol II has been thought to resemble the small single subunit polymerases in two respects, the involvement of two Mg ions in the mechanism of catalysis and the occurrence of an α helix adjacent to the A site, termed the bridge helix in the case of pol II and the O helix in the single subunit enzymes (Temiakov et al., 2003). The first of these similarities is, however, only incidental, and the second is illusory. Pol II and the family of large,

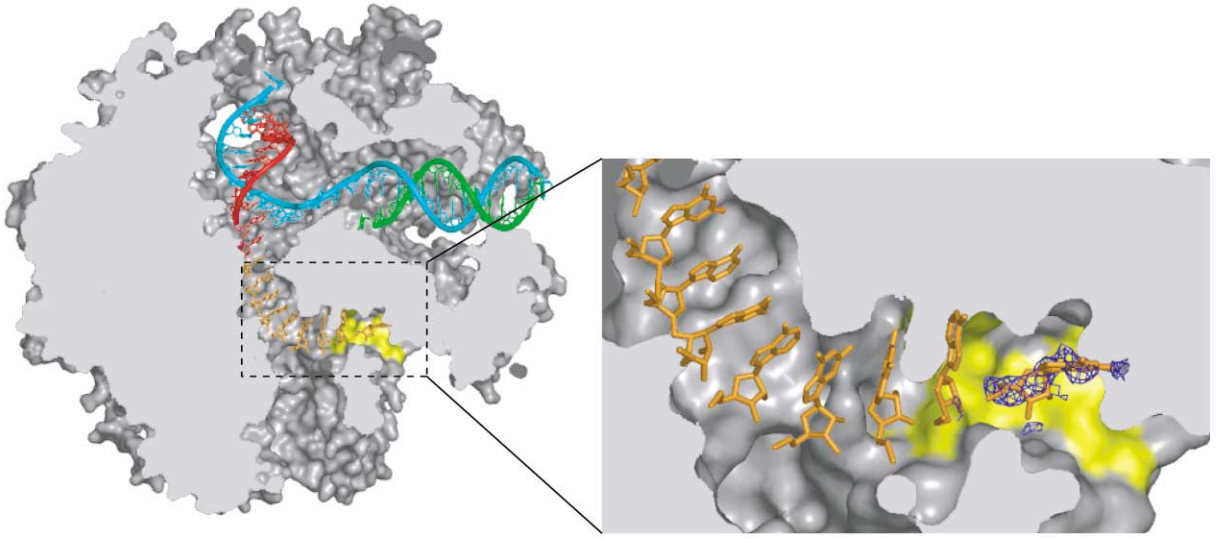

Figure 4. Third Site of NTP Binding to Pol II and Model of a Backtracked Complex

A cutaway of the solvent-accessible surface of a transcribing complex, viewed in the same direction as Figure 1, with RNA in red, template DNA strand in cyan, and nontemplate strand in green. Backtracked RNA is modeled in orange, extending to a binding pocket (yellow surface) where persistent density ($2F_o-F_c$ map in blue) is observed with all NTP crystal soaks.

multisubunit polymerases differ fundamentally from the small single subunit enzymes.

The role of Mg ions in the various catalytic activities of the polymerases may be viewed, like the involvement of Mg ions in the nucleic acid transactions of many other enzymes, as a consequence of the association of Mg^{2+} with nucleotides in their various forms, and especially with NTPs, in all of nature. It is doubtless for this reason that a Mg ion enters with the NTP at the E site. The single and multisubunit polymerases differ even in this aspect, as both Mg ions enter with the NTP in the single subunit enzymes, whereas one Mg ion is permanently resident at the A site of the multisubunit enzymes. Mobility of metal A in T7 RNA polymerase has been suggested to play a role in discriminating between rNTPs and dNTPs (Temiakov et al., 2004), whereas the location of

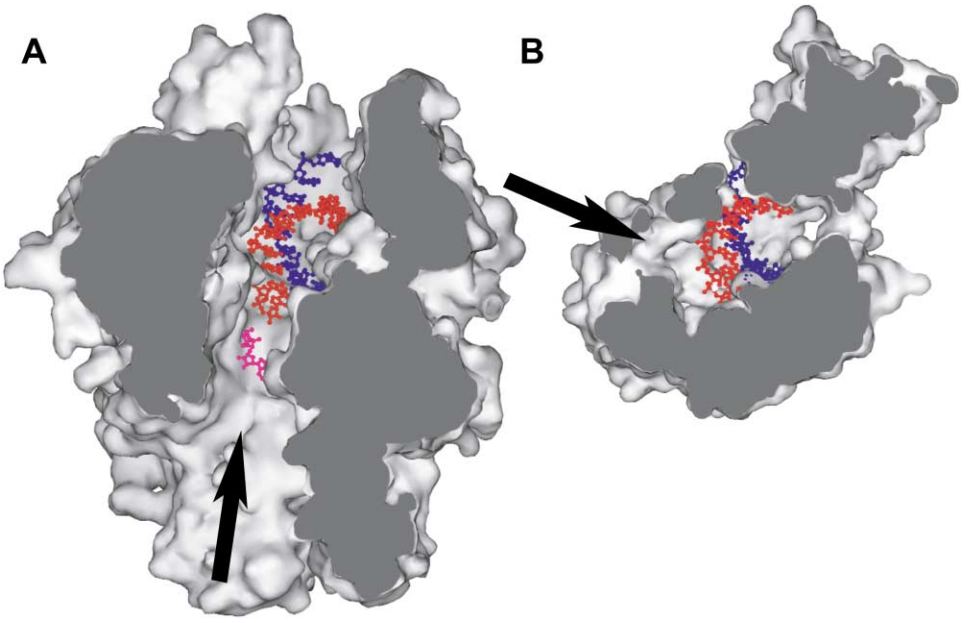

Figure 5. Substrate Entry to Active Center Regions of Single and Multisubunit Polymerases

Solvent-accessible surfaces for transcribing complexes of (A) pol II and (B) T7 RNA polymerase (PDB 1H38) are shown, in a "front" view of pol II (Cramer et al., 2000) and corresponding view of the T7 enzyme (aligned on the sugar-phosphate backbone of the DNA-RNA hybrid) (Tahirov et al., 2002), with the front portion of the proteins cut away to reveal the DNA-RNA hybrid (DNA blue, RNA red). A mismatched NTP bound to pol II as in Figure 2D is shown in pink. The direction of substrate entry to the active center is indicated by a black arrow.

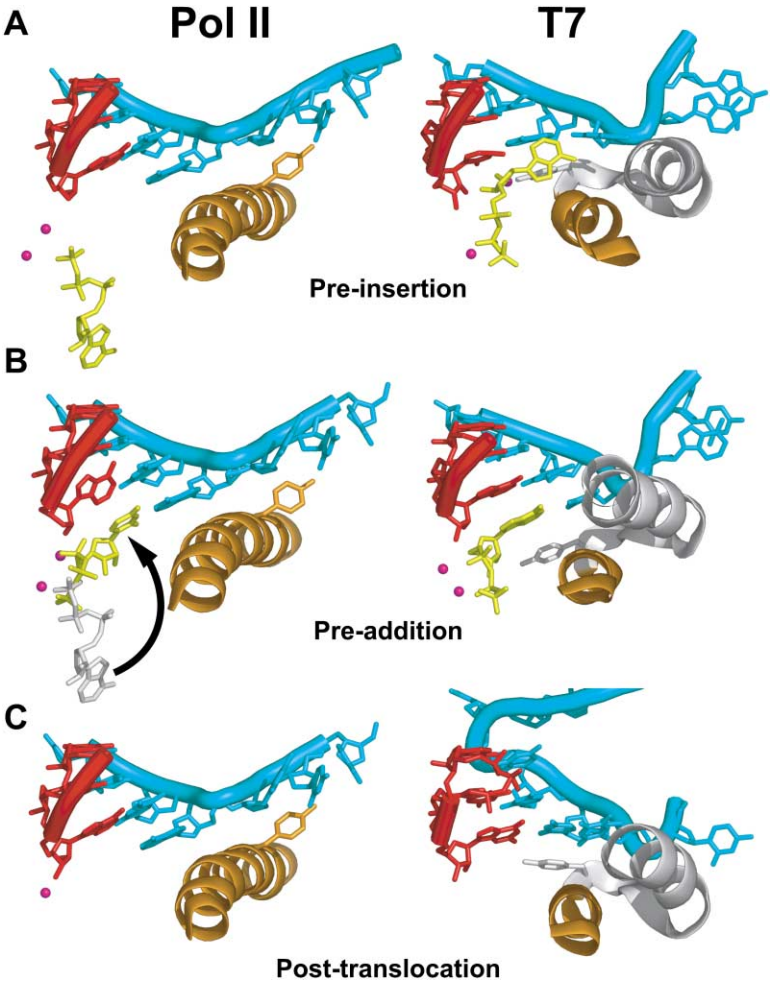

Figure 6. Side-by-Side Comparison of T7 and RNA Polymerase II Transcribing Complex Structures

Nucleic acids near the active centers of pol II (left column) and T7 RNA polymerase II transcribing complexes (right column) (Temiakov et al., 2004; Yin and Steitz, 2004) are shown, with RNA in red, DNA template strand in cyan, the pol II bridge helix and tyrosine residue Y836 in brown, the T7 O helix in brown, the loop, tyrosine residue Y639, and O′ helix that follow the O helix in gray, the incoming NTP in yellow, and Mg ions as magenta spheres. (Top row) Complexes with NTPs before insertion in the nucleotide addition site. (Middle row) Complexes with NTPs following insertion in the nucleotide addition site. The rotation of the NTP in the pol II complex is indicated by an arrow, with the nucleotide before insertion shown in gray. (Bottom row) Complexes following nucleotide addition and translocation.

metal A is invariant in the structures of multisubunit RNA polymerases.

Recent X-ray structures of T7 RNA polymerase transcribing complexes (Temiakov et al., 2004; Yin and Steitz, 2004) permit direct comparison with the pol II transcribing complex. Structures of the pre-translocation complex before and after entry of NTP, and of the posttranslocation complex, have now been determined for both enzymes. It is immediately apparent that the NTP entry pathways differ markedly (Figure 5B). Whereas entry to the pol II active center is along the axis of the DNA-RNA hybrid helix, through a long narrow pore, entry to the T7 active center is perpendicular to the hybrid helix axis, exposed near the surface of the enzyme. For a more detailed comparison of the transcribing complex structures, they were aligned on the sugar-phosphate backbone of the RNA-DNA hybrid (Figure 6). The pol II and T7 structures are then seen to differ at each stage of the transcription process. The paths of the template DNA strand are remarkably dissimilar. The interacting protein elements are unrelated as well. In particular, the bridge helix and O helix are seen to perform very different roles. Whereas the bridge helix contacts the template strand, and especially the base at position i+1, the O helix contacts only the entering NTP, while a loop following the O helix contacts the template. A tyrosine residue, Tyr639, in the loop following the O helix is important for the selection of an NTP rather than dNTP (Brieba and Sousa, 2000; Temiakov et al., 2004; Yin and Steitz, 2004). Y639 is also important for coupling O helix movement to translocation (Yin and Steitz, 2004). The pol II bridge helix contains a tyrosine residue (Tyr836) as well, but it is too far from the incoming NTP to interact. A change from straight to bent conformation of the bridge helix, such as may accompany translocation, rotates Y836 even further away from the NTP.

The difference between the pol II and T7 complexes is also evident in the mechanism of NTP selection by the template DNA base. As already mentioned, NTPs enter beneath the active center of pol II, bind to the E site, and rotate into the A site for pairing with the template base, held fixed by interaction with the bridge helix. In contrast, NTPs enter the T7 complex from the side and bind to a pre-insertion site in the correct orientation for catalysis, requiring only translation and not rotation for binding to the nucleotide addition site. The NTP is held fixed and the template base "flips" into position for base pairing in this site (Patel et al., 2001).

Finally, the mechanism of DNA double helix unwinding in advance of the active center differs between the pol II and T7 complexes. Our structures show that DNA strand separation in the pol II complex is mediated by

interactions between the sugar-phosphate backbone and basic residues in the floor of the active center cleft, whereas strand separation in the T7 complex involves a 22° rotation of a five-helix bundle (including the O helix), with stacking of a phenylalanine residue, F644, on template bases (Yin and Steitz, 2004). Although, as already mentioned, strand separation during transcription may begin further upstream than observed in our structures, the interactions we have identified are likely to be involved and are clearly different than those seen in the T7 complex.

To summarize, the multisubunit RNA polymerases, typified by pol II, differ fundamentally from single subunit RNA and DNA polymerases. There is no sequence conservation, no structural homology, and only coincidental mechanistic similarity between the two systems. They represent separate evolutionary solutions to the problem of nucleic acid polymerization. In the case of pol II, the elaborate architecture of the pore and NTP entry pathway, absent from the single subunit enzymes, enables such associated processes as backtracking and transcript cleavage, important for pausing and regulation, proofreading, and DNA repair.

Experimental Procedures

Crystallization

Pol II was purified as described (Cramer et al., 2000). Transcribing complexes were formed by association of synthetic oligonucleotides with ten-subunit pol II as described (Kireeva et al., 2000). DNA and RNA oligonucleotides were mixed and annealed by raising the temperature to 60°C and gradually decreasing it to 25°C. Pol II and 3′ dATP were added and incubated for 1 hr at room temperature. The final concentrations in the mixture were 20 mM template DNA strand, 20 mM nontemplate DNA strand, 40 mM RNA, 2 mM pol II, 20 mM 3′ dATP, 20 mM Tris, pH 7.5, 40 mM KCl, and 5 mM DTT. Excess oligonucleotides were removed by ultrafiltration. The complexes were crystallized by vapor diffusion as described (Gnatt et al., 2001). Crystals were transferred to freezing buffer and flash frozen in liquid nitrogen as described (Cramer et al., 2001). For the introduction of nucleotides, 15 mM nucleotide and 5 mM $MgCl_2$ were added to the freezing buffer. For the collection of Mn anomalous data, $MnCl_2$ was substituted for $MgCl_2$. Binary complexes of nucleotides with pol II alone were produced by the addition of 15 mM nucleotide and 50 mM $MnCl_2$ to crystals of pol II formed as described (Cramer et al., 2001), prior to freezing.

Data Collection and Processing

Diffraction data were collected at beamlines 9-2 and 11-1 at the Stanford Synchrotron Radiation Laboratory and beamlines 5.0.2 and 8.2.2 at the Advanced Light Source, Lawrence Berkeley National Laboratory. Data were processed in DENZO and SCALEPACK (Otwinowski and Minor, 1996) or MOSFLM (Leslie, 1992) and SCALA (4., 1994). Model building was done in the program O (Jones et al., 1991) and refinement was done using REFMAC (Murshudov et al., 1999) or CNS (Brunger et al., 1998). Modeling of the nucleic acid backbone was based on the nucleic acid from PDB 1I6H, using phosphate peaks as the primary criteria for backbone location. Additional residues were added into unbiased electron density.

Acknowledgments

K.D.W. was supported by the Medical Scientist Training Program. This research was supported by NIH grants GM49985 and AI21144 to R.D.K.

Received: May 19, 2004
Revised: July 31, 2004
Accepted: September 14, 2004
Published: November 11, 2004

References

Artsimovitch, I., Patlan, V., Sekine, S., Vassylyeva, M.N., Hosaka, T., Ochi, K., Yokoyama, S., and Vassylyev, D.G. (2004). Structural basis for transcription regulation by alarmone ppGpp. Cell *117*, 299–310.

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

Brieba, L.G., and Sousa, R. (2000). Roles of histidine 784 and tyrosine 639 in ribose discrimination by T7 RNA polymerase. Biochemistry *39*, 919–923.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science *288*, 640–649.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863–1876.

Doublie, S., Tabor, S., Long, A.M., Richardson, C.C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. Nature *391*, 251–258.

Dudev, T., and Lim, C. (2003). Principles governing Mg, Ca, and Zn binding and selectivity in proteins. Chem. Rev. *103*, 773–788.

Epshtein, V., Mustaev, A., Markovtsov, V., Bereshchenko, O., Nikiforov, V., and Goldfarb, A. (2002). Swing-gate model of nucleotide entry into the RNA polymerase active center. Mol. Cell *10*, 623–634.

Fish, R.N., and Kane, C.M. (2002). Promoting elongation with transcript cleavage stimulatory factors. Biochim. Biophys. Acta *1577*, 287–307.

Foster, J.E., Holmes, S.F., and Erie, D.A. (2001). Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. Cell *106*, 243–252.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science *292*, 1876–1882.

Gu, W., and Reines, D. (1995). Variation in the size of nascent RNA cleavage products as a function of transcript length and elongation competence. J. Biol. Chem. *270*, 30441–30447.

Huang, Y., Eckstein, F., Padilla, R., and Sousa, R. (1997). Mechanism of ribose 2′-group discrimination by an RNA polymerase. Biochemistry *36*, 8231–8242.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47 (Pt 2)*, 110–119.

Kettenberger, H., Armache, K.J., and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. Cell *114*, 347–357.

Kireeva, M.L., Komissarova, N., Waugh, D.S., and Kashlev, M. (2000). The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. J. Biol. Chem. *275*, 6530–6536.

Leslie, A. (1992). Joint CCP4 and ESF-EACMB Newsletter, No. 26 (Warrington, United Kindgom: Daresbury Laboratory).

Murshudov, G.N, Lebedev, A., Vagin, A.A., Wilson, K.S., and Dodson, E.J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr. *D55*, 247–255.

Opalka, N., Chlenov, M., Chacon, P., Rice, W.J., Wriggers, W., and Darst, S.A. (2003). Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. Cell *114*, 335–345.

Otwinowski, Z., and Minor, W. (1996). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Patel, P.H., Suzuki, M., Adman, E., Shinkai, A., and Loeb, L.A. (2001).

Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection. J. Mol. Biol. *308*, 823–837.

Santangelo, T.J., and Roberts, J.W. (2004). Forward translocation is the natural pathway of RNA release at an intrinsic terminator. Mol. Cell *14*, 117–126.

Shi, Y.B., Gamper, H., Van Houten, B., and Hearst, J.E. (1988). Interaction of Escherichia coli RNA polymerase with DNA in an elongation complex arrested at a specific psoralen crosslink site. J. Mol. Biol. *199*, 277–293.

Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V., and Goldfarb, A. (2003). Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. EMBO J. *22*, 2234–2244.

Steitz, T.A. (1998). A mechanism for all polymerases. Nature *391*, 231–232.

Tahirov, T.H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W.T., Vassylyev, D.G., and Yokoyama, S. (2002). Structure of a T7 RNA polymerase elongation complex at 2.9 A resolution. Nature *420*, 43–50.

Temiakov, D., Tahirov, T.H., Anikin, M., McAllister, W.T., Vassylyev, D.G., and Yokoyama, S. (2003). Crystallization and preliminary crystallographic analysis of T7 RNA polymerase elongation complex. Acta Crystallogr. D Biol. Crystallogr. *59*, 185–187.

Temiakov, D., Patlan, V., Anikin, M., McAllister, W.T., Yokoyama, S., and Vassylyev, D.G. (2004). Structural basis for substrate selection by t7 RNA polymerase. Cell *116*, 381–391.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. Science *303*, 1014–1016.

Yin, Y.W., and Steitz, T.A. (2004). The structural mechanism of translocation and helicase activity in t7 RNA polymerase. Cell *116*, 393–404.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. Cell *98*, 811–824.

Accession Numbers

Structure factors and coordinates for pol II elongation complexes and pol II-NTP binary complexes have been deposited at the Protein Data Bank with ID codes 1R9T, 1R9S, 1TWF, 1TWC, 1TWA, 1TWG, and 1TWH.