Cell, Vol. 116, 381–391, February 6, 2004, Copyright ©2004 by Cell Press

Structural Basis for Substrate Selection by T7 RNA Polymerase

Dmitry Temiakov,[1,2] Vsevolod Patlan,[3] Michael Anikin,[1,2] William T. McAllister,[1] Shigeyuki Yokoyama,[2,3,4,5,*] and Dmitry G. Vassylyev[2,3,*]

[1]Morse Institute for Molecular Genetics
Department of Microbiology
SUNY Health Science Center
450 Clarkson Avenue
Brooklyn, New York 11203

[2]Cellular Signaling Laboratory

[3]Structurome Research Group
RIKEN Harima Institute at Spring-8
1-1-1 Kouto, Mikazuki-cho, Sayo
Hyogo 679-5148
Japan

[4]RIKEN Genomic Sciences Center
1-7-22 Suehiro-cho
Tsurumi, Yokohama 230-0045
Japan

[5]Department of Biophysics and Biochemistry
Graduate School of Science
University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-0033
Japan

Summary

The mechanism by which nucleotide polymerases select the correct substrate is of fundamental importance to the fidelity of DNA replication and transcription. During the nucleotide addition cycle, pol I DNA polymerases undergo the transition from a catalytically inactive "open" to an active "closed" conformation. All known determinants of substrate selection are associated with the "closed" state. To elucidate if this mechanism is conserved in homologous single subunit RNA polymerases (RNAPs), we have determined the structure of T7 RNAP elongation complex with the incoming substrate analog. Surprisingly, the substrate specifically binds to RNAP in the "open" conformation, where it is base paired with the acceptor template base, while Tyr639 provides discrimination of ribose versus deoxyribose substrates. The structure therefore suggests a novel mechanism, in which the substrate selection occurs prior to the isomerization to the catalytically active conformation. Modeling of multisubunit RNAPs suggests that this mechanism might be universal for all RNAPs.

Introduction

The single subunit RNA polymerase encoded by bacteriophage T7 (T7 RNAP; 98 kDa) is structurally related to the pol I family of DNA polymerases (DNAPs) yet carries out all of the steps in the transcription process in a nearly identical manner to that observed for multisubunit RNAPs (Jeruzalmi and Steitz, 1998; McAllister, 1997; McAllister and Raskin, 1993; Sousa et al., 1993). It therefore provides a useful model for understanding fundamental aspects of nucleotide polymerization in the context of RNA synthesis. Unlike DNAPs, which extend a preexisting primer, RNAPs must initiate transcription de novo by incorporating the two ribonucleoside triphosphates that are directed by the sequence at the start site of the promoter. After it clears the promoter, the RNAP isomerizes to form a stable, highly processive elongation complex (EC). Another distinguishing feature of transcription (as opposed to DNA replication) is that the RNAP must displace the nascent transcript from the template strand. During elongation, the two strands of DNA are separated at the leading edge of transcription complex and reannealled at the trailing edge, forming a transcription "bubble" that encloses an 8 bp RNA:DNA hybrid (Huang and Sousa, 2000; Temiakov et al., 2000).

The C-terminal domain of T7 RNAP exhibits the same overall architecture as the pol I family of DNAPs, resembling a cupped right hand with fingers, palm, and thumb domains (Jeruzalmi and Steitz, 1998; Sousa et al., 1993). The structure around the active site in the palm domain is highly conserved among T7 RNAP and DNAPs and contains two aspartate residues that coordinate the two metal ions that are required during the phosphoryl transfer reaction (Jeruzalmi and Steitz, 1998; Steitz et al., 1994; Steitz and Steitz, 1993). The fingers domain contains the highly conserved O and O′ α helices (Y and Z helices in T7 RNAP), which in DNAPs were shown to interact with the incoming deoxy NTP (dNTP) substrate and the template base to which it is paired (Figure 1). To simplify the comparison between DNAPs and RNAPs, we refer to Y and Z helices in RNAP as to O and O′ helices, respectively.

In both RNAPs and DNAPs, each cycle of nucleotide addition can be divided into several distinct steps (Figure 1A) (Yin and Steitz, 2004 [this issue of *Cell*]). Structural studies of pol I DNAPs have identified two major states (catalytically active and inactive) of the enzyme (Figure 1) (Doublie and Ellenberger, 1998; Johnson et al., 2003; Kiefer et al., 1998; Li et al., 1998b). In the catalytically active "closed" conformation, the O/O′ helices are proximal to the active center, and the substrate is bound by interactions with a number of conserved basic residues in the O helix (Figure 1B). In this configuration, the substrate dNTP and the acceptor DNA nucleotide at position n (i+1 in RNAP nomenclature) of the template strand (TS_n) are located in the "insertion" site, where the formation of a correct Watson-Crick base pair is required to achieve catalysis, thereby preventing incorporation of mismatched bases (Figure 1B) (Doublie and Ellenberger, 1998; Johnson et al., 2003). The discrimination against ribonucleotides (rNTPs) also occurs in the insertion site through a "steric gate," in which the substrate deoxyribose is sandwiched between the side chains of invariant Glu and Phe residues (Astatke et al., 1998; Joyce, 1997).

*Correspondence: yokoyama@biochem.s.u-tokyo.ac.jp (S.Y.), dmitry@yumiyoshi.harima.riken.go.jp (D.G.V.)

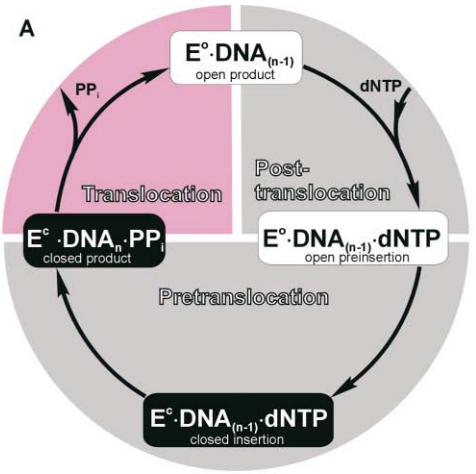
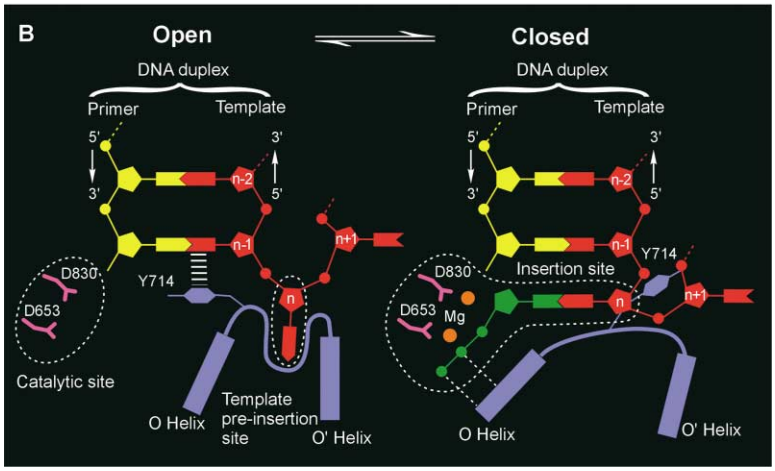


Figure 1. Nucleotide Addition Cycle

(A) Schematic drawing reflecting four possible states of the protein/nucleic acid complex during nucleotide addition cycle. E°, E°–open and closed conformations of the enzyme, respectively; n, (n-1)—positions occupied by the 3′-terminal primer nucleotides incorporated in the DNA.

(B) Two conformational states of DNAPs during the nucleotide addition cycle. The figure is based on the schematic drawing previously done for *B. stearothermophilus* enzyme (Johnson et al., 2003); the numbering of active site aspartates and the tyrosine blocking the entrance of the insertion site for the TS_n base in the "open" conformation corresponds to that of *B. stearothermophilus* DNAP.



The catalytically inactive state is characterized by an "open" conformation in which the O/O′ helices are distal from the catalytic center, and the portion of the insertion site that will later be occupied by the TS_n base is blocked by a Tyr residue from the tip of the O helix (Y639 in the O helix of T7 RNAP) (Figure 1B). The inactive configuration forms a "preinsertion" binding site in which the TS_n base is sequestered in a protein pocket formed by O/O′ helices (Figure 1B) (Johnson et al., 2003). So far, no structural data are available for substrate bound with high affinity and specificity in the preinsertion site of ternary DNAP complexes (Johnson et al., 2003; Kiefer et al., 1998; Li et al., 1998a, 1998b).

The transition from an open to closed configuration is accompanied by movement of helices O and O′, which results in displacement of the blocking Tyr residue (Figure 1B). Judging by the high level of structural homology with DNAPs in the vicinity of the active site, a similar transition between open and closed forms is likely to occur in T7 RNAP. The state of RNAP in the ternary RNAP elongation complex (EC), which corresponds to the open conformation of DNAPs, has been referred to as "semi-open," as more extensive opening of the enzyme is observed in the absence of nucleic acids (Yin and Steitz, 2004). To allow for an easier comparison between RNAPs and DNAPs, we will use the term "open" in place of "semi-open" throughout this paper.

Previously reported structures of T7 RNAP initiation (Cheetham and Steitz, 1999) and elongation complexes (Tahirov et al., 2002; Yin and Steitz, 2002) were obtained in the absence of substrate and correspond to an open conformation in which the TS_n base is in the preinsertion position (Huang and Sousa, 2000; Ma et al., 2002). A number of residues that affect the fidelity of base incorporation (Gly640, Phe644, Gly645, His784) or provide discrimination against incorporation of dNTPs (Tyr639) have been identified in T7 RNAP (Brieba and Sousa, 2000; Huang et al., 2000; Sousa and Padilla, 1995). When mapped onto the existing structures of T7 RNAP, these residues (except for Tyr639) are located far from the catalytic center, making it difficult to interpret these data by structural analogy with DNAPs. The mechanism by which T7 RNAP achieves substrate selection and its coupling with the catalytic reaction was therefore obscure.

In this work, we determined the structure of a T7 RNAP EC with an analog of the incoming substrate ribonucleotide (rNTP) specifically bound in the preinsertion site. Significantly, the rNTP is observed to interact with the acceptor DNA base, allowing the "fit" of the incipient

base pair to be sampled prior to formation of the catalytically active closed conformation. The structure suggests that discrimination against dNTPs also commences while substrate is bound in the preinsertion site. Inspection of the structures of multisubunit bacterial RNAPs suggests that similar mechanisms for substrate selection may exist in these enzymes as well.

Results and Discussion

Structure Determination and Overall Structure

To elucidate the mode of substrate binding, T7 RNAP EC were assembled on a nucleic acid scaffold that directs the incorporation of ATP downstream of an 8 bp hybrid (Tahirov et al., 2002; Temiakov et al., 2002, 2003) and were cocrystallized with the nonhydrolyzable ATP analog (α,β-methylene ATP, AMPcPP), resulting in crystals that diffracted beyond 3 Å. Fourier maps with coefficients of |2Fobs − Fcalc|, calculated after initial rigid body and B factor refinement for the data, showed clear electron density (ED) for the protein residues and additional ED for the ribose and tri-phosphate moieties of AMPcPP in all four independent molecules in the asymmetric unit. A Mg^{2+} ion bound to the β and γ phosphates of AMPcPP could also be fit in the ED. However, significant ED for the adenine base of the AMPcPP was observed in only two of four molecules at this initial stage of refinement. Atomic models for AMPcPPs were built into the ED and refinement was carried out until it converged to a final R factor of 25.5% (Rfree = 31.1%) at 3.2 Å resolution. The refinement substantially improved the quality of the model and phases, so that ED was clearly visible for the base moieties of the AMPcPP in all four molecules, as revealed by a final simulated annealing omit |Fobs − Fcalc| ED map (Figure 2A). However, significant residual ED was present in close vicinity to the ribose O2′ atoms of the AMPcPPs in all four RNAP molecules. Very similar strong EDs were also present at the same positions in the initial ED maps. Water molecules were introduced into the atomic coordinates and the refinement was carried out from the beginning. Surprisingly, despite the strong repulsive interactions that should prevent a close approach of these water molecules and the AMPcPP O2′ oxygens, the refinement brought the water and O2′ oxygens to a distance as short as ~2.2–2.3 Å, which corresponds to the actual centers of the peaks in the omit ED maps. The most likely interpretation for this result was that the residual ED reflects the presence of Mg^{2+} ions that are coordinated by the O2′ oxygens of the AMPcPP and Tyr639 hydroxyls. Consistent with this, refinement carried out with Mg^{2+} atoms in these positions resulted in ~8–9σ level peaks in the final simulated annealing omit |Fobs − Fcalc| ED maps (Figure 2A).

The overall structure of the RNAP EC in complex with the substrate analog is very similar to that observed previously in the EC in the absence of substrate, with root mean square deviation of about ~1.2 Å overall main chain atoms and corresponds to the open configuration in which the TS_n base remains in the preinsertion position (Figure 2B) (Tahirov et al., 2002; Yin and Steitz, 2002). The four independent RNAP EC structures in the asymmetric unit of the crystal are highly similar. However, it is worth noting that the TS_n DNA base which occupies the preinsertion site is located at slightly different distances from the active site in the four different independent RNAP molecules in the crystal.

Substrate Binding Site

In the present structure, the substrate is bound along the open O helix at a distance of ~10 Å from the active center in a nearly identical manner in all four molecules in the asymmetric unit of the crystal (Figures 2B and 2C). The phosphates of the AMPcPP are fixed through hydrogen bonds with three basic residues, Arg627, Lys631, Arg632 in helix O and Lys472 and Tyr571, all of which are well conserved (Figure 2C). Substitutions of all these residues have dramatic effects on NTP binding and enzyme activity (Figure 3A) (Bonner et al., 1992, 1994; Guajardo and Sousa, 1997; Huang et al., 2000; Osumi-Davis et al., 1992; Tunitskaya and Kochetkov, 2002; Woody et al., 1998). Interactions of NTP with two of these five residues (Arg627 and Lys631) persist in the closed substrate complex (Yin and Steitz, 2004); thus, although the modes of substrate binding in the preinsertion and insertion sites are different, there is a partial overlap between these sites by virtue of common interacting residues in the O helix. These observations indicate that substrate binding in the open state as observed here is specific and that the complex is a functionally important intermediate in the nucleotide addition cycle.

Significantly, the base moiety of AMPcPP in the T7 RNAP EC-NTP complex is located within interacting distance from the base of the complementary DNA TS_n thymine in the preinsertion sites in all four RNAP molecules (Figure 2C). Though the substrate and DNA template nucleotides do not form exact Watson-Crick base pairs, their relative orientations are appropriate to achieve base-specific recognition in the preinsertion site upon subtle movements of the TS_n base and/or substrate. Consistent with this observation, the best ED for the substrate base is observed in the RNAP molecule in which the distance between the substrate and complementary DNA bases is the shortest (~3.1 Å). In another molecule in which the substrate base is located further away from the complementary DNA base (~4.2 Å) its position is less well defined.

An important characteristic of the substrate bound in the preinsertion site is the presence of a Mg^{2+} ion bound to the O2′ of the AMPcPP ribose, which is clearly observed in all four RNAP molecules. This Mg^{2+} ion is coordinated by the hydroxyl of Tyr639, which blocks entry of the TS_n nucleotide of the DNA into the insertion site and stacks upon the template base at n-1 (Figure 2C). The Mg^{2+}-mediated interaction between the Y639 OH and the 2′OH of the ribose accounts for the key role of Y639 in the discrimination of the ribo-versus deoxyribo nucleotides (Brieba and Sousa, 2000; Huang et al., 1997b; Kostyuk et al., 1995; Sousa and Padilla, 1995). This Mg^{2+} ion might be also coordinated by His784, the carbonyl oxygen of Met635, and interestingly, by the ribose O2′ atom of the 3′ rNMP of the RNA/DNA hybrid (n-1 position) through water-mediated interactions. The substrate ribose and base moieties are additionally fixed through stacking interactions with the side chain of Met635, consistent with previous biochemical studies (Figure 3A) (Guajardo and Sousa, 1997).

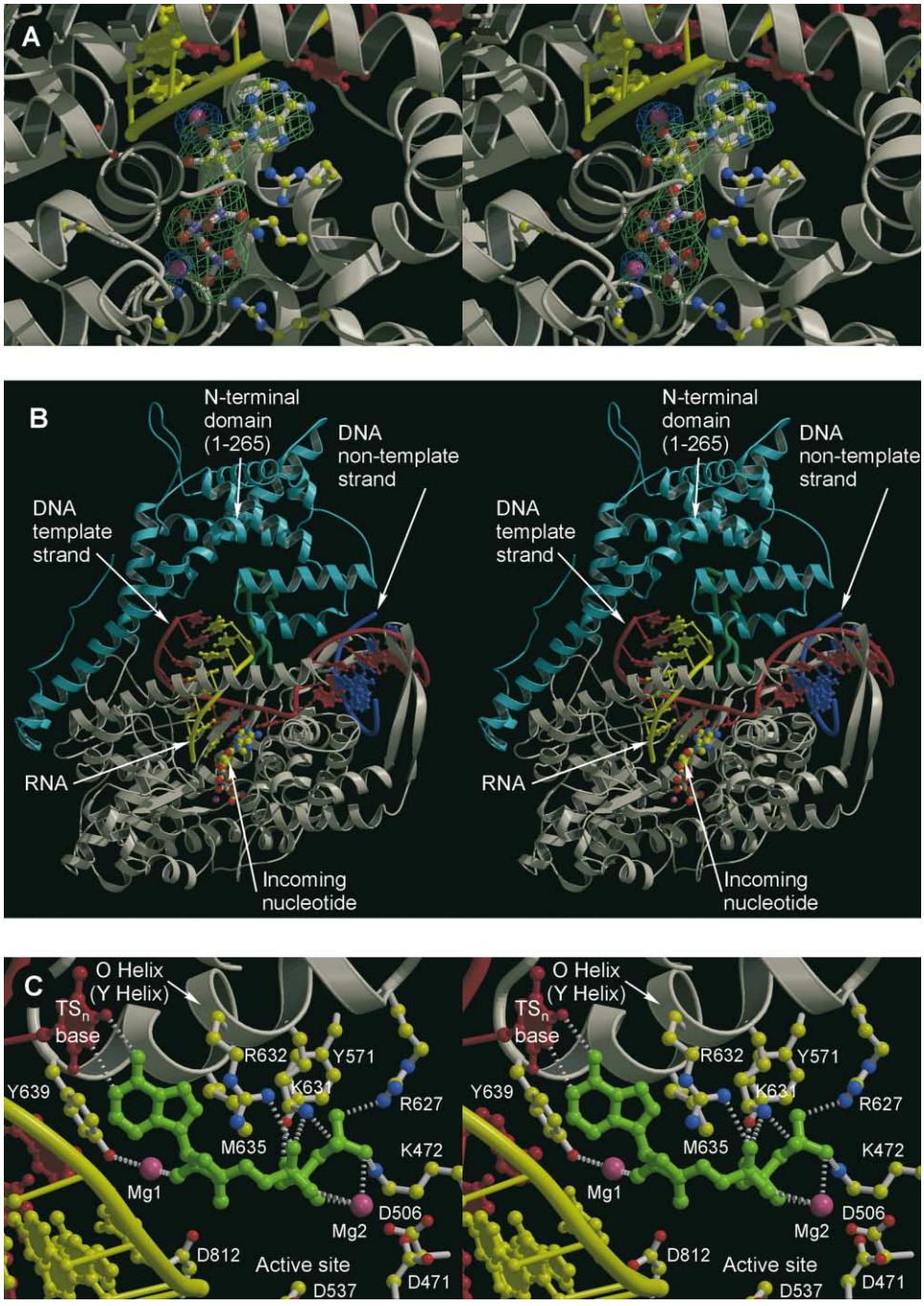


Figure 2. Structure of T7 RNAP EC with Bound AMPcPP

The protein structure is shown by ribbon diagram. The refolded N-terminal domain and the rest of the protein are cyan and white, respectively. RNA, DNA template, and nontemplate nucleotides are shown in yellow, red, and blue, respectively. AMPcPP (green sticks) and protein side chains (white sticks) involved in a substrate binding are shown by ball-and-stick models. Mg^{2+} ions are shown as magenta spheres.

(A) Simulated annealing |Fobs − Fcalc| omit electron density maps produced after refinement for substrate (green, 3σ level) and Mg^{2+} ions (blue, 5σ level).

(B) Overall view of the T7 RNAP EC with AMPcPP bound in the substrate entry channel along the open O helix.

(C) Substrate binding site. Hydrogen bonds and Mg^{2+} coordination bonds are shown as white dashed sticks.

Another Mg^{2+} ion is bound to the β and γ phosphates of the AMPcPP (Figure 2C). Although there are two conserved acidic residues, Asp471 and Asp506, in the vicinity of the phosphate bound Mg^{2+} ion (Figure 2C), these residues seem to be too far away to interact directly with the metal ion, but potentially might be involved in Mg^{2+} coordination through water-mediated interactions. The binding mode of Mg^{2+} to the phosphates of the AMPcPP is quite similar to that of one of the catalytic Mg^{2+} ions in DNAP (Doublie et al., 1998) and RNAP (Yin

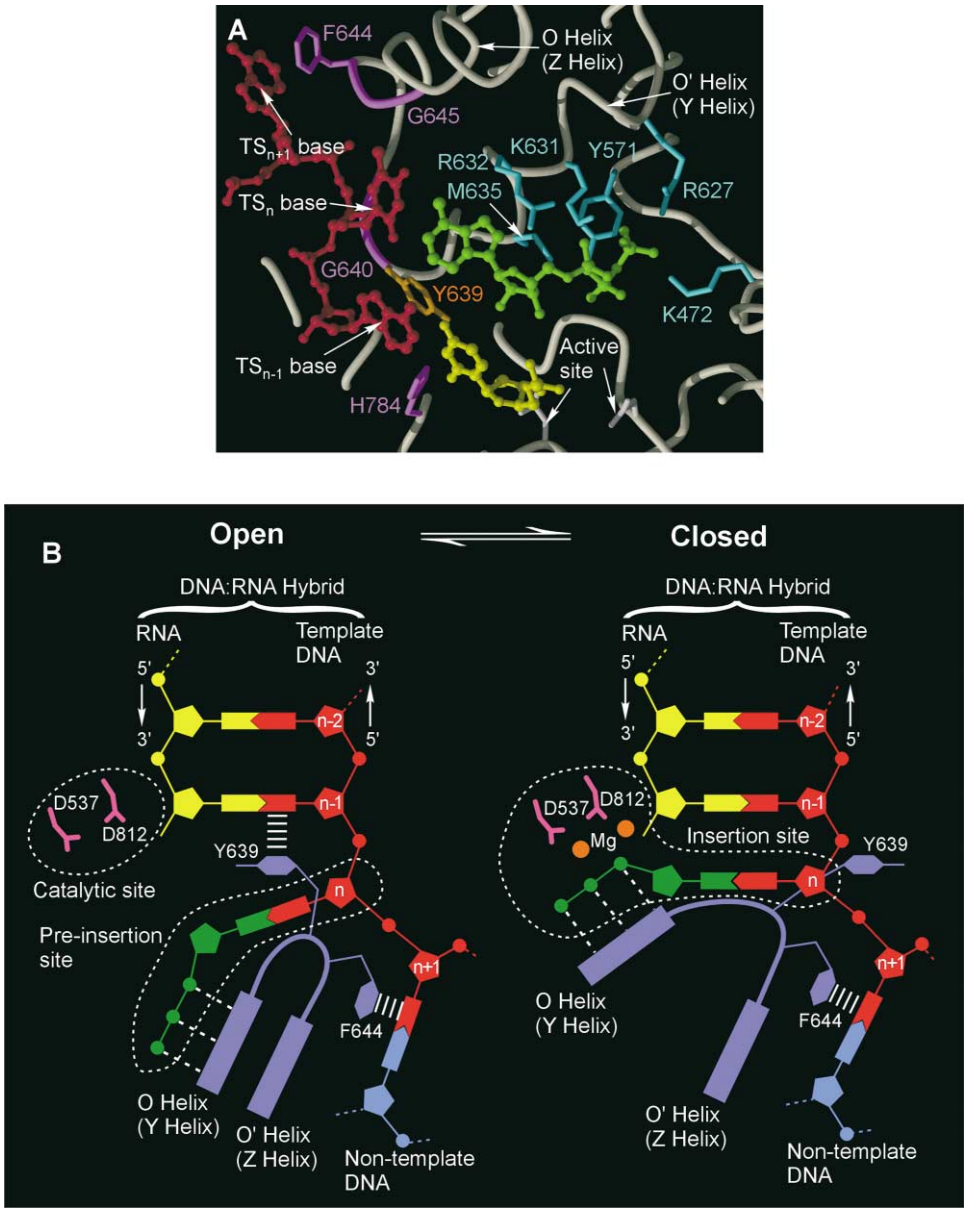


Figure 3. Mechanism of Substrate Selection by T7 RNAP

(A) Mapping of mutations in the substrate binding site. The residues, mutations of which affect substrate binding, incorporation of mismatched bases, or dNTPs are shown in cyan, magenta, and orange, respectively. The RNA, DNA template nucleotides and substrate are shown in yellow, red, and green, respectively.

(B) Schematic drawing of the two RNAP conformations corresponding to substrate (green) selection and catalysis during RNA synthesis.

and Steitz, 2004) closed complexes, and it is possible that this ion would move to the active site together with the phosphates of the substrate.

Substrate Selection by T7 RNAP

The structure provides two major implications. First, it shows the pathway through which NTP substrate gains access to the active site of the enzyme, as binding of the substrate to the closed insertion site is sterically impossible (Yin and Steitz, 2004). Second, the binding mode of AMPcPP in the preinsertion site indicates that, at least in part, this site is used to achieve selection of the proper substrate prior to its movement into the catalytic center (insertion site) (Figure 3B). Indeed, the two major criteria for substrate selection, recognition of the O2′ ribose atom and proper base pairing with the complementary DNA nucleotide, are likely to commence in this site.

According to the present structural results, the orientations of both substrate and TS$_n$ nucleotide in the preinsertion site should be crucial for the fidelity of base incorporation. The base of the TS$_n$ nucleotide has no direct interactions with the protein and is located in a large protein cavity (Figure 4A), nevertheless its orientation is fixed by constraints imposed by the phosphate backbone that links the nucleotide to the adjacent up-

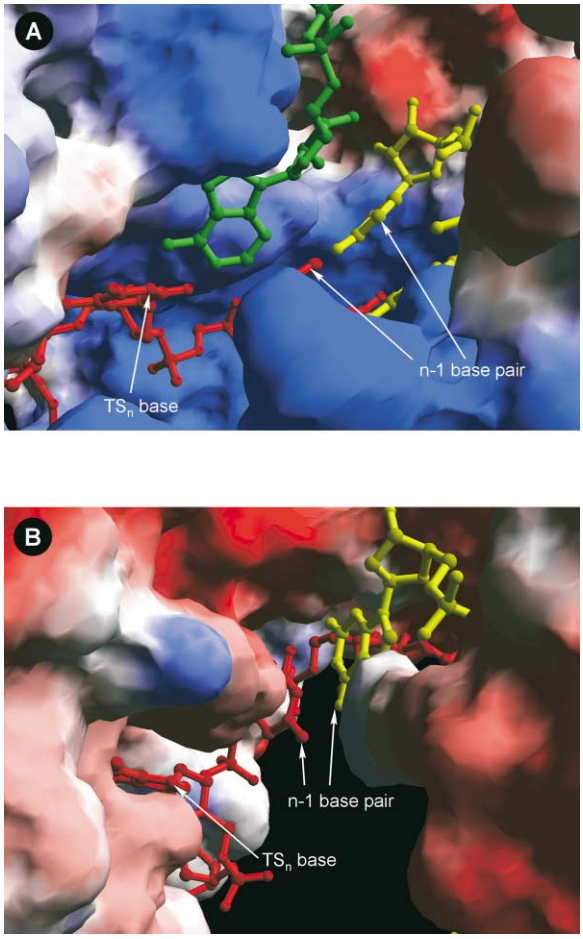


Figure 4. Structural Organization of Preinsertion Sites

(A) T7 RNAP. (B) *B. stearothermophilus* DNAP. The proteins are represented by the surfaces of electrostatic potentials (red, blue, and white are positively charged, negatively charged, and neutral regions, respectively).

stream and downstream T strand nucleotides (TS_{n-1}, TS_{n+1}) (Figure 3A). Consistent with the proposed role of the preinsertion site in substrate selection, three of the four T7 RNAP mutants that are known to exhibit a significant increase in incorporation of mismatched bases (G640A, F644A, G645A) (Huang et al., 2000) are involved in the loop that intervenes between the O and O′ helices (which defines a portion of the preinsertion site) and interact with TS_{n+1} base and TS_n/TS_{n+1} phosphate backbone (Figure 3A). In addition, substitution of His784 (H784A), which is within interacting distance from the TS_{n-1} nucleotide and may therefore affect the orientation of the TS_n base, results in both misincorporation errors and extension of mismatched base pairs (Huang et al., 2000) (Figure 3A).

The structure indicates that selection of rNTPs versus dNTPs could also occur in the preinsertion site in T7 RNAP through Mg^{2+}-mediated interaction of the Tyr639 hydroxyl with the 2′OH group of substrate ribose. This is in an excellent agreement with the role of Tyr639 previously revealed by biochemical and genetic results (Figure 3A) (Brieba and Sousa, 2000; Huang et al., 1997a, 1997b; Kostyuk et al., 1995; Sousa and Padilla, 1995). The Y639F mutation, which results in a 20-fold increase in dNTP incorporation, is of particular interest, as it accounts for the primary role of the Tyr hydroxyl group in rNTP selection.

In the closed complex (Yin and Steitz, 2004), the substrate retains base-specific and ribose-discriminating interactions that might prove crucial for the fidelity of RNA synthesis beyond the preliminary substrate selection observed in the preinsertion site. However, it seems unlikely that substrate-specific interactions would be disrupted after their formation in the open complex and then restored in the closed configuration. We therefore suggest that, once established in the preinsertion site, these interactions would be maintained or even enhanced during transition from the open to the closed states. The requirement of a close fit of the active site to a properly formed base pair in order for catalysis to occur (Doublie et al., 1998) would thus provide a final sieve in terms of the fidelity of substrate incorporation (Figure 3B) (Yin and Steitz, 2004).

Comparison with DNAPs

In DNAPs, the fidelity of nucleotide incorporation is believed to rely largely on the close fit of the active site (insertion site) in the closed complex to a strictly complementary Watson-Crick base pair, such that the geometry of an incorrectly formed base pair would not be consistent with catalysis (Doublie et al., 1998; Johnson et al., 2003). This fit is ensured in part by a "steric gate," which plays a major role in selection of the deoxyribo substrates (Astatke et al., 1998; Joyce, 1997). The structure of DNAP indicates that, as in RNAP, once the closed conformation has been achieved, the insertion site is not accessible for binding of exogenous substrate. It is therefore likely that substrate binding must occur when the enzyme is in the open configuration. There are two possible pathways for substrate binding in the open DNAP conformation.

First, the substrate might bind the O helix mostly through its phosphates as observed in T7 RNAP and in the two binary substrate complexes of DNAPs (Kiefer et al., 1998; Li et al., 1998b). In principle, in this mode of binding, initial substrate selection might occur in a manner similar to that of T7 RNAP. However, an important difference between T7 RNAP and DNAPs concerns the manner in which the TS_n nucleotide is positioned in the preinsertion site. In contrast to T7 RNAP (discussed above) (Figures 3B and 4A), the TS_n base in DNAPs is sequestered in a tight pocket formed by the O and O′ helices (Figures 1 and 4B), thereby precluding potential interactions with the substrate if it would bind along the open O helix in a similar manner to that of T7 RNAP (Johnson et al., 2003). Another important structural consideration arguing against the close similarity of substrate selection mechanisms in the open state between DNAPs and RNAP is that the "steric gate" controlling selection of dNTPs in DNAPs can be entirely formed only in the closed conformation, whereas no other mechanism which may substitute for the "steric gate" in the open form is evident. Finally, in DNAPs, all known substitutions that affect incorporation of mismatches and/or discrimination of dNTPs versus rNTPs are con-

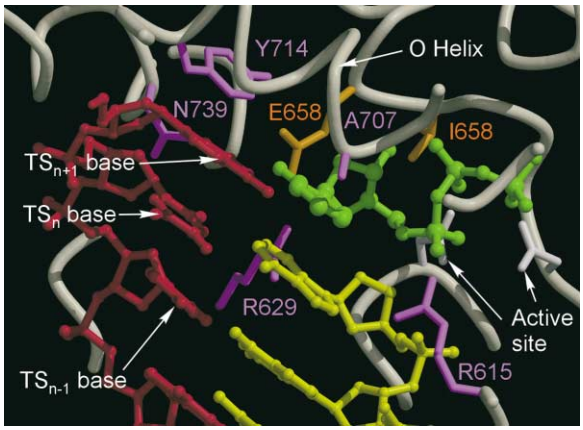


Figure 5. Mapping of Substitutions on the Active Site Region of *Bacillus stearothermophilus* DNAP

The residues, substitutions of which affect incorporation of mismatched bases, or both mismatched bases and rNTPs are shown in magenta and orange, respectively. The DNA template, primer, and substrate are shown in red, yellow, and green, respectively.

centrated in the vicinity of the insertion site (Figure 5) (Minnick et al., 1999, 2002; Patel et al., 2001; Shinkai and Loeb, 2001; Tosaka et al., 2001). Though most of these residues have no direct interactions with the substrate, they are likely either to reduce the space in the active site or to destabilize the closed conformation in the presence of mismatched bases (Minnick et al., 1999; Patel et al., 2001), thereby precluding formation of improper base pairs between the substrate and the TS_n nucleotide in the closed insertion site (Figure 5). On the other hand, in the structure of *B. stearothermophilus* DNAP (PDB ID 1L3S), Arg789 interacts with the phosphate backbone of the TS_{n-1} nucleotide and stacks on the TS_{n+1} base in the preinsertion site; this residue might therefore affect substrate miscincoporation in an analogous manner as residues in the preinsertion site in T7 RNAP if these enzymes utilize a similar mechanism of substrate selection. However, substitution of its counterpart in *E. coli* DNAP (Arg849) had no effect on the fidelity of substrate incorporation (Minnick et al., 1999). Nevertheless, we cannot rule out the possibility that initial base pairing between substrate and the TS_n nucleotide may occur in DNAP during the transition from the open to the closed conformation, as movement of the O helix may expose the TS_n base as it is released from the preinsertion site (Johnson et al., 2003). This would be consistent with the observation that substitution of a conserved tyrosine in DNAPs, which is analogous to Tyr639 of T7 RNAP, substantially increased incorporation of mismatched base pairs, though it did not affect discrimination against rNTPs (Bell et al., 1997; Minnick et al., 1999). This tyrosine could also assist rNTP biding by DNAPs, but only when steric gating residue (Glu710 in *E. coli* DNAP) is truncated (E710A) to eliminate its competing interactions with the tyrosine side chain (Astatke et al., 1998; Brieba and Sousa, 2000). This latter observation leaves open the possibility that DNAPs and RNAPs discriminate rNTPs versus dNTPs by not altogether different mechanisms.

In the second mode of binding, the substrate may bind directly to the insertion (catalytic) site in the open configuration, in which case it is likely to be fixed mostly through its base (via interactions with theTyr639 counterpart) and deoxyribose moieties, as has been observed in a ternary DNAP complex (Li et al., 1998b). However, the latter complex was obtained in the crystal of a closed ternary complex upon dissociation of the substrate from the insertion site and is therefore somewhat artificial. Moreover, as interactions of the substrate phosphates with basic residues in the O helix would be lacking (as compared to the mode of binding described above), the affinity of the substrate is likely to be rather low.

To conclude, the structural data available thus far are insufficient to favor one of these pathways in DNAPs, and the possibility of substrate binding/selection in the open form of DNAP remains to be tested.

Multisubunit RNAPs

Recently determined structures of the T7 RNAP EC revealed that the organizational changes that occur during the transition from initiation to the elongation phase of transcription result in an unexpected similarity at the structural level between the single and multisubunit enzymes (Cramer et al., 2001; Gnatt et al., 2001; Tahirov et al., 2002; Vassylyev et al., 2002; Yin and Steitz, 2002). All previously identified structural elements that are characteristic of the stable ECs in the multisubunit enzymes (substrate entry and RNA exit channels, RNA/DNA hybrid binding cavity, and downstream DNA binding sites) are also present in the T7 RNAP EC. Given this, it is reasonable to expect a certain level of similarity in fundamental aspects of transcription including substrate binding and discrimination. To evaluate the possibility that multisubunit enzymes might utilize a mechanism of substrate selection similar to that observed in T7 RNAP, we constructed a model of the *T. thermophilus* RNAP docked with the RNA/DNA hybrid and downstream DNA from the T7 EC (in which the TS_n acceptor base is in the preinsertion position).

It had been suggested that in multisubunit RNAPs, the bridge helix is involved in DNA translocation and the delivery of the TS_n DNA nucleotide to the active site (Epshtein et al., 2002; Gnatt et al., 2001; Murakami and Darst, 2003; Tahirov et al., 2002; Vassylyev et al., 2002). This helix would therefore be the functional equivalent of the O-O′ helices of the single subunit DNAPs (RNAPs). The two states of Tyr639 in the T7 RNAP O helix are likely to control the movement of the incoming DNA template nucleotide between the preinsertion and insertion sites (Yin and Steitz, 2004). Similarly, two states of the bridge helix have been observed in the mulitsubunit RNAPs. Whereas in yeast RNAP the helix is uniform, in the *T. thermophilus* RNAP, two residues in the helix are flipped out and the central portion of the helix is distorted (Cramer et al., 2001; Gnatt et al., 2001; Vassylyev et al., 2002). The yeast RNAP transcription complex is in the pretranslocated state, in which the base pair formed by the newly incorporated substrate and the TS_n base is still in the active site; the position of the TS_n base in this structure would therefore correspond to the TS_n base in the insertion site in the closed conformation of T7 RNAP (c.f., Figures 4B and 6A, right panels). The

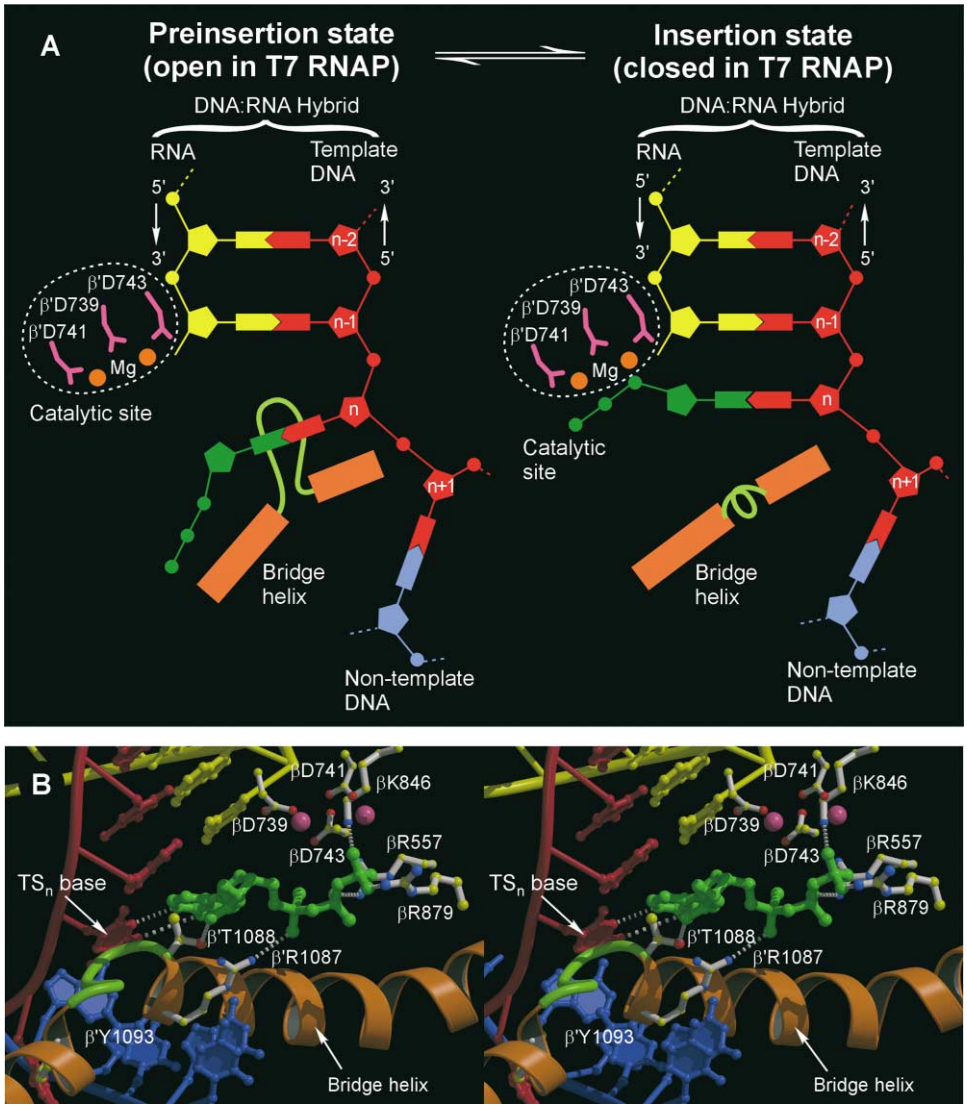


Figure 6. Putative Mechanism of Substrate Selection by Multisubunit RNAPs

(A) Schematic drawing of two possible conformations (distorted and uniform) of the bridge helix in multisubunit RNAPs corresponding to the preinsertion (distorted) and insertion (uniform) sites for the TS_n nucleotide and substrate. The region of bridge helix that is distorted in the preinsertion site is shown in light green; the substrate is green. Sequence numbering corresponds to that of *T. thermophilus* RNAP.

(B) Stereo view of the RNA:DNA hybrid of the T7 RNAP EC and substrate base paired with the DNA TS_n nucleotide in the preinsertion site modeled into the *T. thermophilus* RNAP structure (distorted conformation of the bridge helix).

distorted conformation of the helix in the *T. thermophilus* RNAP would thus correspond to the T7 RNAP EC in the open configuration (c.f., Figures 3B and Figure 6A, left panels). Modeling of the RNA/DNA hybrid from the T7 RNAP EC into the *T. thermophilus* RNAP structure reveals that the DNA TS_n nucleotide in the preinsertion conformation fits nicely to the distorted conformation of the bridge helix (Figure 6B). Modeling of the substrate base paired with the DNA acceptor TS_n base in the preinsertion position does not result in clash of the substrate with the protein and suggests that the substrate phosphates, although lacking interactions with active site Mg^{2+} ions, might be fixed in this preinsertion mode by the same triad (Argβ557, Argβ879, Lysβ846) of basic residues that are likely to bind the substrate in the active site. Argβ'1087 from the bridge helix, which has been demonstrated to form a crosslink with the 3′ RNA nucleotide in the RNA:DNA hybrid (Epshtein et al., 2002) is likely to bind to the substrate α phosphate and/or ribose, thereby increasing the affinity of the substrate for the preinsertion site.

Recently, based on kinetic and mutational analyses, the presence of an allosteric substrate binding site was proposed for multisubunit RNAPs, and it was suggested that presampling of substrates prior to their delivery to the catalytic center might occur in this site (Foster et al., 2001; Holmes and Erie, 2003). Interestingly, the proposed substrate "prebinding" site is in close proximity to the preinsertion site described in our work (though not identical).

Concluding Remarks

The structure of the incoming nucleotide bound to the T7 RNAP EC in the open conformation, together with

Table 1. Data Collection and Refinement Statistics

Space group		P1	
Unit cell parameters (Å)		a = 81.13, b = 87.70, c = 206.53	
(°)		α = 91.93, β = 91.02, γ = 110.66	
Resolution (Å)		40.0–3.0 (3.11–3.0)[a]	
Reflections (Total/Unique)		229,915/94,902	
I/σ(I)		7.1 (2.3)	
R_{merge} (%)		8.4 (46.1)	
Completeness (%)		89.4 (83.9)	
Refinement		**Model quality**	
Resolution (Å)	40.0–3.2 (3.31–3.2)	RMSD bond length (Å)	0.022
Reflections used	73,174 (7,731)	RMSD bond angles (°)	2.1
R_{factor} (%)	25.5 (37.7)	RMSD improper angles (°)	1.46
R_{free} (%)	30.7 (41.6)		
Number of protein atoms	26,984		
Number of nucleic acid atoms	2,802		
Number of AMPcPP atoms	124		
Number of Mg^{2+} atoms	8		
Number of water molecules[b]	999		

$R_{merge} = \Sigma_{hkl}\Sigma_j \mid I_j(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_j \langle I(hkl)\rangle$, where $I_j(hkl)$ and $\langle I(hkl)\rangle$ are the intensity of measurement *j* and the mean intensity for the reflection with indices *hkl*, respectively. $R_{factor,\,free} = \Sigma_{hkl}||F_{calc}(hkl)| - |F_{obs}(hkl)||/\Sigma_{hkl}|F_{obs}|$, where the crystallographic R factor is calculated including and excluding reflections in the refinement. The free reflections constituted 4% of the total number of reflections. RMSD—root mean square deviation. I/σ(I)—ratio of mean intensity to a mean standard deviation of intensity.

[a] The data for the highest resolution shell are shown in brackets.

[b] The water molecules with electron density peaks less than 1.25σ in the $|2F_{obs} - F_{calc}|$ electron density map were removed from the model.

the previously determined structures of the posttranslocation complex (Tahirov et al., 2002; Yin and Steitz, 2002) and two structures of the closed ternary RNAP complexes (Yin and Steitz, 2004), provide a complete picture of all four steps in the nucleotide polymerization cycle (Figure 1A). It also provides insights into the mechanism by which the incoming NTP substrate gains access to the active site of RNAP, as well as the structural basis for substrate selection.

However, two important questions remain open. The first is whether the base-specific and ribose-discriminating interactions with the substrate established in the open ternary complex are the principal determinants of the fidelity of substrate incorporation or whether they correspond to initial contacts that require further proofreading in the closed state. Additional biochemical analyses are required to address this question. We anticipate that substitutions of residues other than Gly640, Phe644, Gly645 near the preinsertion site, as well as those around the insertion site (similar to analyses performed for DNAPs), may shed light on the mechanism of substrate selection by T7 RNAP.

Another important question is whether DNAPs also utilize a mechanism involving substrate selection in the open state as observed for T7 RNAP. As discussed above, both structural considerations and available biochemical data argue against the existence of such a mechanism. However, more data will be required to draw a final conclusion. Substitutions in the vicinity of the preinsertion site, for example in the loop intervening between O and O′ helices, may help to elucidate the mode of substrate selection by DNAPs.

Experimental Procedures

Crystallization and Data Collection

T7 RNAP elongation complexes were assembled on a nucleic acid scaffold as previously described (Tahirov et al., 2002, 2003). The sequences of the nucleic acid components were: T strand: GGGAATCGATATCGCCGC; NT strand: ATCGATTCCC; RNA: AACU GCGGCGAU. Crystallization was carried out under conditions described before (Temiakov et al., 2003) in the presence of α,β-methylene ATP (4 mM, Sigma). Large, cube-like crystals (0.4 × 0.4 × 0.4 mm) were obtained by microseeding against 7% PEG 8000, 6% MPD in 100 mM Tris (pH 7.9). Crystals were flash-frozen in a cryoprotectant solution containing 8% PEG 8000, 6% MPD, 20% ethylene glycol, 50 mM NaCl, 10 mM $MgCl_2$ in 50 mM Tris (pH 7.9). Diffraction data were collected at 3.0 Å resolution using synchrotron radiation at Spring-8 beam line BL26 and processed using the HKL2000 program package (Otwinowski and Minor, 1997).

Structure Determination and Refinement

Though the space group of the crystals was the same as for the EC, the unit cell parameters were substantially changed (Table 1) (Tahirov et al., 2002, 2003). Therefore, the initial R factor of the EC model was ~45% at 3.2 Å resolution. The crystals were characterized by the same type of twinning observed previously for the EC crystals (Tahirov et al., 2002, 2003). However, in the present crystals, the twinning was even more prominent. Though crystals diffracted to 2.9 Å resolution, 3.2 Å resolution data produced better quality of the electron density than that obtained for higher resolution data. After rigid body refinement carried out by the CNS program (Brunger et al., 1998), the Rfree dropped to 36.5%. At this stage, the model for AMPcPP was built into the omit |2Fobs − Fc| electron density map in all four RNAP molecules in the asymmetric unit. After several subsequent cycles of positional and B factor refinements using strong noncrystallographic symmetry restraints and alternate "water pick" and "water delete" procedures as implemented in the CNS program (Brunger et al., 1998), refinement converged to the final R factor of 25.5% (Rfree = 31.1%) (Table 1). The relatively high value of Rfree probably accounts for the relatively low resolution of the data used, as well as in part by the presence of twinning, which likely affected the quality of the processed diffraction data. With few exceptions, the EDs in the final |2Fo-Fc| and omit |Fo-Fc| maps were of sufficiently high quality to determine functionally important structural elements (AMPcPP, Mg^{2+} ions, protein residues bound to the substrate and TS_n nucleotide) (Figure 2A) in all four independent molecules in the asymmetric unit of the crystals. However, judging by the modest resolution of the data, the structure might not be precise in terms of detailed structural information (i.e., exact conformations of the NTP phosphates and protein side chains, proper positions, and coordination of Mg^{2+} ions, etc). Figure 5 was prepared using the SWISS-PDB Viewer program (Kaplan and Littlejohn, 2001).

Figures 2, 3, 4A, and 6B were drawn using the MOLSCRIPT (Kraulis, 1991), BOBSCRIPT (Esnouf, 1999), and Raster3D (Merrit and Bacon, 1997) programs. The atomic coordinates for the T7RNAP substrate complex reported here, as well as the corresponding structure factors, have been deposited in the Protein Data Bank with accession number 1S0V.

Acknowledgments

We thank Drs. R. Sousa and I. Artsimovitch for critical reading of the manuscript and helpful discussions during manuscript preparation. This work was supported in part by grant GM38147 from the NIH (USA) to W.T.M. and by the RIKEN Structural Genomics/Proteomics Initiative (RSGI, the National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, Science and Technology of Japan) to S.Y. and D.G.V.

Received: September 17, 2003
Revised: November 12, 2003
Accepted: December 22, 2003
Published: February 5, 2004

References

Astatke, M., Ng, K., Grindley, N.D., and Joyce, C.M. (1998). A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. Proc. Natl. Acad. Sci. USA *95*, 3402–3407.

Bell, J.B., Eckert, K.A., Joyce, C.M., and Kunkel, T.A. (1997). Base miscoding and strand misalignment errors by mutator Klenow polymerases with amino acid substitutions at tyrosine 766 in the O helix of the fingers subdomain. J. Biol. Chem. *272*, 7345–7351.

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

Bonner, G., Lafer, E.M., and Sousa, R. (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. *D54*, 905–921.

Cheetham, G.M., and Steitz, T.A. (1999). Structure of a transcribing T7 RNA polymerase initiation complex. Science *286*, 2305–2309.

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., and Ellenberger, T. (1998). The mechanism of action of T7 DNA polymerase. Curr. Opin. Struct. Biol. *8*, 704–712.

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 Å resolution. Nature *391*, 251–258.

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.

Esnouf, R.M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. Acta Crystallogr. *D55*, 938–940.

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 Å resolution. Science *292*, 1876–1882.

Guajardo, R., and Sousa, R. (1997). A model for the mechanism of polymerase translocation. J. Mol. Biol. *265*, 8–19.

Holmes, S.F., and Erie, D.A. (2003). Downstream DNA sequence effects on transcription elongation. Allosteric binding of nucleoside triphosphates facilitates translocation via a ratchet motion. J. Biol. Chem. *278*, 35597–35608.

Huang, J., and Sousa, R. (2000). T7 RNA polymerase elongation complex structure and movement. J. Mol. Biol. *303*, 347–358.

Huang, J., Brieba, L.G., and Sousa, R. (2000). Misincorporation by wild-type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. Biochemistry *39*, 11571–11580.

Huang, Y., Beaudry, A., McSwiggen, J., and Sousa, R. (1997a). Determinants of ribose specificity in RNA polymerization: effects of Mn2+ and deoxynucleoside monophosphate incorporation into transcripts. Biochemistry *36*, 13718–13728.

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

Jeruzalmi, D., and Steitz, T.A. (1998). Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. EMBO J. *17*, 4101–4113.

Johnson, S.J., Taylor, J.S., and Beese, L.S. (2003). Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. Proc. Natl. Acad. Sci. USA *100*, 3895–3900.

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

Kaplan, W., and Littlejohn, T.G. (2001). Swiss-PDB viewer (deep view). Brief. Bioinform. *2*, 195–197.

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

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

Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Li, Y., Kong, Y., Korolev, S., and Waksman, G. (1998a). Crystal structures of the Klenow fragment of Thermus aquaticus DNA polymerase I complexed with deoxyribonucleoside triphosphates. Protein Sci. *7*, 1116–1123.

Li, Y., Korolev, S., and Waksman, G. (1998b). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. EMBO J. *17*, 7514–7525.

Ma, K., Temiakov, D., Jiang, M., Anikin, M., and McAllister, W.T. (2002). Major conformational changes occur during the transition from an initiation complex to an elongation complex by T7 RNA polymerase. J. Biol. Chem. *277*, 43206–43215.

McAllister, W.T. (1997). Transcription by T7 RNA polymerase. Nucl. Acids Mol. Biol. *11*, 15–25.

McAllister, W.T., and Raskin, C.A. (1993). The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. Mol. Microbiol. *10*, 1–6.

Merrit, E.A., and Bacon, D.J. (1997). Raster3D: photorealistic molecular graphics. Methods Enzymol. *277*, 505–524.

Minnick, D.T., Bebenek, K., Osheroff, W.P., Turner, R.M., Jr., Astatke, M., Liu, L., Kunkel, T.A., and Joyce, C.M. (1999). Side chains that influence fidelity at the polymerase active site of Escherichia coli DNA polymerase I (Klenow fragment). J. Biol. Chem. *274*, 3067–3075.

Minnick, D.T., Liu, L., Grindley, N.D., Kunkel, T.A., and Joyce, C.M. (2002). Discrimination against purine-pyrimidine mispairs in the polymerase active site of DNA polymerase I: a structural explanation. Proc. Natl. Acad. Sci. USA *99*, 1194–1199.

Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the whole story. Curr. Opin. Struct. Biol. *13*, 31–39.

Osumi-Davis, P.A., de Aguilera, M.C., Woody, R.W., and Woody, A.Y. (1992). Asp537, Asp812 are essential and Lys631, His811 are catalytically significant in bacteriophage T7 RNA polymerase activity. J. Mol. Biol. *226*, 37–45.

Otwinowski, Z., and Minor, W. (1997). Processing 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.

Shinkai, A., and Loeb, L.A. (2001). In vivo mutagenesis by Escherichia coli DNA polymerase I. Ile(709) in motif A functions in base selection. J. Biol. Chem. *276*, 46759–46764.

Sousa, R., and Padilla, R. (1995). A mutant T7 RNA polymerase as a DNA polymerase. EMBO J. *14*, 4609–4621.

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

Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. Proc. Natl. Acad. Sci. USA *90*, 6498–6502.

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

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 Å resolution. Nature *420*, 43–50.

Temiakov, D., Anikin, M., and McAllister, W.T. (2002). Characterization of T7 RNA polymerase transcription complexes assembled on nucleic acid scaffolds. J. Biol. Chem. *277*, 47035–47043.

Temiakov, D., Mentesana, P.E., Ma, K., Mustaev, A., Borukhov, S., and McAllister, W.T. (2000). The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. Proc. Natl. Acad. Sci. USA *97*, 14109–14114.

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.

Tosaka, A., Ogawa, M., Yoshida, S., and Suzuki, M. (2001). O-helix mutant T664P of Thermus aquaticus DNA polymerase I: altered catalytic properties for incorporation of incorrect nucleotides but not correct nucleotides. J. Biol. Chem. *276*, 27562–27567.

Tunitskaya, V.L., and Kochetkov, S.N. (2002). Structural-functional analysis of bacteriophage T7 RNA polymerase. Biochemistry (Moscow) *67*, 1124–1135.

Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. Nature *417*, 712–719.

Woody, A.Y., Osumi-Davis, P.A., Hiremath, M.M., and Woody, R.W. (1998). Pre-steady-state and steady-state kinetic studies on transcription initiation catalyzed by T7 RNA polymerase and its active-site mutants K631R and Y639F. Biochemistry *37*, 15958–15964.

Yin, Y.W., and Steitz, T.A. (2002). Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. Science *298*, 1387–1395.

Yin, W.Y., and Steitz, T.A. (2004). The mechanism of translocation and helicase activity in T7 RNA polymerase. Cell *116*, this issue, 393–404.