THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 289, NO. 35, pp. 24549–24559, August 29, 2014

Structural Basis of Transcription Initiation by Bacterial RNA Polymerase Holoenzyme*[S]

Received for publication, May 22, 2014, and in revised form, June 23, 2014 Published, JBC Papers in Press, June 27, 2014, DOI 10.1074/jbc.M114.584037

Ritwika S. Basu[‡], Brittany A. Warner[‡1], Vadim Molodtsov[‡], Danil Pupov[§], Daria Esyunina[§], Carlos Fernández-Tornero[¶], Andrey Kulbachinskiy[§], and Katsuhiko S. Murakami[‡2]

From the [‡]Department of Biochemistry and Molecular Biology, The Center for RNA Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, [¶]Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, 28040 Madrid, Spain, and the [§]Laboratory of Molecular Genetics of Microorganisms, Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia

Background: Cellular RNA polymerases start transcription by *de novo* RNA priming.
Results: Structures and biochemical studies of initially transcribing complexes elucidate the *de novo* transcription initiation and early stage of RNA transcription.
Conclusion: 5′-end of RNA in the transcribing complex starts σ ejection from core enzyme.
Significance: Insights from this study can be applicable to all cellular RNA polymerases.

The bacterial RNA polymerase (RNAP) holoenzyme containing σ factor initiates transcription at specific promoter sites by *de novo* RNA priming, the first step of RNA synthesis where RNAP accepts two initiating ribonucleoside triphosphates (iNTPs) and performs the first phosphodiester bond formation. We present the structure of *de novo* transcription initiation complex that reveals unique contacts of the iNTPs bound at the transcription start site with the template DNA and also with RNAP and demonstrate the importance of these contacts for transcription initiation. To get further insight into the mechanism of RNA priming, we determined the structure of initially transcribing complex of RNAP holoenzyme with 6-mer RNA, obtained by *in crystallo* transcription approach. The structure highlights RNAP-RNA contacts that stabilize the short RNA transcript in the active site and demonstrates that the RNA 5′-end displaces σ region 3.2 from its position near the active site, which likely plays a key role in σ ejection during the initiation-to-elongation transition. Given the structural conservation of the RNAP active site, the mechanism of *de novo* RNA priming appears to be conserved in all cellular RNAPs.

During successive steps of transcription initiation, RNAP[3] must employ different mechanisms of NTP loading and accommodate the growing RNA product within the RNA channel until entering the stable transcription elongation phase. All cellular and most bacteriophage RNAPs initiate transcription *de novo* by loading two iNTPs opposite the first and second template DNA bases at the transcription start site to synthesize the first phosphodiester bond of RNA. Loading the first iNTP, which will become the 5′-end of RNA, at the *i* site is a unique process for *de novo* transcription by RNAP because this site in the transcription elongation complex usually accommodates the RNA 3′-end, whose binding is stabilized by the preceding ~8 base pairs of the DNA/RNA hybrid (1). The binding affinity of the first iNTP during *de novo* transcription is substantially lower than that of the second iNTP binding at the *i*+1 site (2, 3). This allows the direct sensing of NTP concentrations by RNAP to become a basis of regulating transcription initiation at ribosomal RNA promoters (4) and as well at the pyrimidine biosynthesis genes (5) in bacteria. High-resolution crystal structures of the *de novo* transcription initiation complex containing RNAP, DNA, and the first and second iNTPs bound at the active site have been determined for bacteriophage T7 (6) and N4 (7) RNAPs, and these structures revealed unique interactions between the first iNTP and RNAP/DNA. However, due to structural differences between different classes of RNAPs, insights from the bacteriophage RNAP structures cannot be directly transferred to cellular RNAPs. Previously published crystal structure of the *Thermus thermophilus* RNAP transcription initiation complex contains a GpA dinucleotide primer complementary to the template DNA positions −1 and +1 but lacking the 5′-triphosphate group (8). Recently, a structure of the *T. thermophilus* RNAP initiating complex containing two iNTPs has been reported (9). However, the roles of the observed RNAP-iNTP contacts in transcription initiation were not tested experimentally; in addition, the structure contained a suboptimal template strand sequence around the transcription start site (see below), suggesting that it might miss some important contacts with the iNTPs.

Following the *de novo* incorporation step, RNAP goes through several cycles of NTP addition before entering transcription elongation, during which a highly stable and proces-

* This work was supported, in whole or in part, by National Institutes of Health Grant GM087350-A1 (to K. S. M.). This work was also supported in part by the Spanish government through Grant BFU2010-16336 (to C. F.-T.). This work was also supported by the Russian Foundation for Basic Research (Grants 14-04-01696 and 14-04-31994) and the Russian Academy of Sciences Presidium Program in Molecular and Cellular Biology (to A. K.).
The atomic coordinates and structure factors (codes 4Q5S and 4Q4Z) have been deposited in the Protein Data Bank (http://wwpdb.org/).
[S] This article contains supplemental Movie S1.
[1] Present address: Janssen Research and Development, L.L.C., Spring House, PA 19477.
[2] To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802. Tel.: 814-865-2758; E-mail: kum14@psu.edu.
[3] The abbreviations used are: RNAP, RNA polymerase; iNTP, initiating ribonucleoside triphosphate; Pol II, RNA polymerase II.

sive RNAP complex performs synthesis of thousands of bases of RNA. In contrast, the initially transcribing complexes of RNAP holoenzyme containing short RNAs, usually ranging 2–12 nucleotides in length, are unstable resulting in abortive initiation (10, 11). Multiple events during initial transcription, including σ release from the core enzyme, DNA scrunching, and promoter escape, have been proposed (12–14); however, molecular basis of the initial transcription by bacterial RNAP is lacking due to the limited stability of these complexes and the difficulty of capturing a homogeneous initiating complex containing short nascent RNA.

Here, we report a crystal structure of the *de novo* transcription initiation complex of *T. thermophilus* RNAP, containing two iNTPs loaded in the active site. The structure and also structure-based biochemical assays reveal key interactions between the RNAP holoenzyme, DNA, and iNTPs that are critical for *de novo* transcription initiation. Furthermore, we prepared a homogeneous initially transcribing complex with a 6-mer RNA using *in crystallo* transcription approach and solved its crystal structure, which provides insights into the binding of short RNA-DNA hybrid in the active site and in the process of σ release triggered by the nascent RNA, as the initially transcribing complex begins transition into the elongation phase of transcription.

EXPERIMENTAL PROCEDURES

Preparation and Crystallization of the T. thermophilus RNAP Promoter DNA Complex, the de Novo Transcription Initiation Complex, and the Transcription Complex Containing 6-Mer RNA—*T. thermophilus* HB8 cells were cultured by using a 300-liter BioService fermentor at the Penn State fermentation facility. Endogenous *T. thermophilus* RNAP core enzyme was purified as follows: ~100 g of cell paste was suspended in 300 ml of lysis buffer (40 mM Tris-HCl, pH 8, at 4 °C, 100 mM NaCl, 10 mM EDTA, 10 mM 2-mercaptoethanol, 1 mM benzamidine, 1 mM PMSF, 0.5 μg/ml leupeptin, and 0.1 μg/ml pepstatin), and cells were lysed by Emulsiflex C3 homogenizer (Avestin, Inc.) at 20,000 psi. After 30 min, benzamidine and PMSF (1 mM) were added to the lysate. The lysate was then clarified by centrifugation, and glycerol was added to the supernatant to a concentration of 5%. RNAP in the soluble fraction was precipitated by adding 10% polyethyleneimine (Polymin-P) solution (0.5%), and the pellet was recovered by centrifugation. The pellet was then resuspended and washed with 200 ml of wash buffer (40 mM Tris-HCl, pH 8, at 4 °C, 200 mM NaCl, 1 mM EDTA, 10 mM 2-mercaptoethanol, 5% glycerol, 1 mM benzamidine, and 1 mM PMSF) and again recovered by centrifugation. This wash step was repeated with an additional 200 ml of wash buffer. Finally, RNAP was recovered by resuspending the pellet in 100 ml of extraction buffer (40 mM Tris-HCl, pH 8, at 4 °C, 1 M NaCl, 1 mM EDTA, 10 mM 2-mercaptoethanol, 5% glycerol, 1 mM benzamidine, and 1 mM PMSF) and centrifuging for 15 min at 4 °C and 17,000 rpm. This extraction step was repeated with another 100 ml of extraction buffer, and the supernatants from both extractions were combined (total, 200 ml). Ammonium sulfate powder was then gradually added to 45% concentration to precipitate RNAP followed by centrifugation. The pellet was suspended in TGED buffer (10 mM Tris-HCl, pH 8 at 4 °C, 0.1 mM EDTA, 5% glycerol, and 1 mM DTT), and RNAP core enzyme was purified by heparin, ResourceQ, and SP Sepharose column chromatography (GE Healthcare).

T. thermophilus σ^A was expressed in BL21(DE3) cells transformed with a pET21a plasmid containing *T. thermophilus sigA* gene. Cells were grown in 1.5 L LB containing 100 μg/ml ampicillin for 6 h, induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside, and grown for additional 3 h at 37 °C. After cells were lysed by sonication, σ^A was purified from the lysate by heat treatment in a 65 °C water bath for 30 min. The suspension was centrifuged at 16,000 rpm to remove the white precipitate. σ^A was then purified from the supernatant by ResourceQ column (GE Healthcare) chromatography. σ^A was stored in 50 mM Tris-HCl, pH 7.7, 400 mM NaCl, and 15% glycerol at −20 °C.

T. thermophilus RNAP holoenzyme was prepared by adding 3-fold molar excess of σ^A to core RNAP and then purified by Superdex200 column chromatography. The RNAP and promoter DNA complex was prepared by mixing 18 μM (24 μl) *T. thermophilus* holoenzyme (in 20 mM Tris-HCl, pH 7.7, 100 mM NaCl, and 1% glycerol) and 1 mM (0.65 μl) of the DNA scaffold (see Fig. 1*A*) and incubated for 30 min at 22 °C. Crystals were obtained by using hanging drop vapor diffusion by mixing equal volume of RNAP-DNA complex solution and crystallization solution (100 mM Tris-HCl, pH 8.7, 200 mM KCl, 50 mM $MgCl_2$, 10 mM Spermine tetra-HCl, and 10% PEG 4000) and incubating at 22 °C over the same crystallization solution. The crystals were cryoprotected by soaking in same constituents as the crystallization solution with stepwise increments of PEG4000 and (2*R*,3*R*)-(−)-2,3-butanediol (Sigma-Aldrich) to final concentrations of 25 and 15%, respectively. The final cryoprotectant solution was also used for soaking NTPs at room temperature to prepare the *de novo* transcription initiation complex (5 mM ATP; and 5 mM CMPCPP from Jena Bioscience; soaking time, 30 min) and the holoenzyme transcription complex with 6-mer RNA (5 mM each of ATP, CTP, and UTP; soaking time, 5 h).

X-ray Data Collections and Structure Determination—The crystals belong to the C-centered monoclinic space group (Table 1) containing one *T. thermophilus* RNAP transcription complex per asymmetric unit. The data set was collected at the Macromolecular Diffraction at Cornell High Energy Synchrotron Source (MacCHESS) F1 beamline (Cornell University, Ithaca, NY), and the data were processed by HKL2000 (15). The *T. thermophilus* RNAP-promoter DNA complex (8) was used as a search model for the molecular replacement (16). Rigid body refinements were performed, and further adjustments to the model were performed manually. The resulting model phases allowed positioning of iNTPs in the *de novo* transcription initiation complex and RNA in the initially transcribing complex in their electron-density maps. Positional refinement and reference model restraints was performed by the program Phenix (17).

In Vitro Transcription—The wild-type *Escherichia coli* RNAP core enzyme and holoenzymes containing σ^{70} were prepared as described (18). Mutations in the *E. coli rpoB* gene of the RNAP β subunit were obtained by site-directed mutagenesis of plasmid pIA545 and recloned into plasmid pIA679 encoding all four core RNAP subunits with a His_6 tag in the N termi-

TABLE 1
Data collection and refinement statistics
Data sets were collected at MacCHESS-F1 (Ithaca, NY).

Complex	*De novo*[a]	ITC[b]
PDB code	4Q4Z	4Q5S
Data collection		
Space group	C2	C2
Cell dimensions	a = 184.49, b = 102.16, and c = 294.72 Å; β=98.96°	a = 187.09, b = 102.08, and c = 297.26 Å; β=98.14°
Resolution (Å)	50–2.85	50–3.00
Total reflections	260,182	273,453
Unique reflections	111,819	107,824
Redundancy	2.3 (1.4)[c]	2.5 (1.9)[c]
Completeness (%)	88.3 (67.7)[c]	94.4 (78.8)[c]
I/σ	5.4 (1.4)[c]	6.6 (1.2)[c]
R_{sym}	0.144 (0.508)[c]	0.124 (0.719)[c]
Refinement		
Resolution (Å)	50–2.90	50–3.00
R_{work}	0.256	0.269
R_{free}	0.275	0.294
r.m.s.d.[d]		
Bond length	0.005 Å	0.002 Å
Bond angles	0.938°	0.61°

[a] The *de novo* transcription complex of RNAP holoenzyme.
[b] The initially transcribing complex of RNAP holoenzyme containing a 6-mer RNA.
[c] Highest resolution shells are shown in parentheses.
[d] r.m.s.d., root mean square deviation.

nus of the β subunit (both plasmids kindly provided by I. Artsimovitch). Mutant RNAPs were expressed in *E. coli* BL21(DE3) and purified using Polymin-P precipitation, heparin, and nickel-nitrilotriacetic acid affinity chromatography, and MonoQ column ion exchange chromatography (19). The *E. coli* RNAP holoenzyme containing the σ subunit with region 3.2 deletion was prepared as described (20).

In vitro transcription of *E. coli* RNAP containing wild-type and mutant β subunits was carried out as described (20). Apparent K_m values for iNTPs were measured as described (20). RNA products were separated by 15% (for full-length RNA transcription), 23% or 30% (for abortive RNAs) PAGE, followed by phosphorimaging.

Analysis of DNA Sequences around Transcription Start Site of Human Pol II Promoters—Human RNAP II (Pol II) promoter sequences (10,000 of 25,976 in database) were obtained from The Eukaryotic Promoter Database (21) and analyzed for their sequence conservations. Graphical representation was prepared using WebLogo (22).

RESULTS

Design of the X-ray Crystallographic Experiment to Determine the Structure of the de novo Transcription Initiation Complex—The *T. thermophilus* RNAP holoenzyme-promoter DNA complex crystals were obtained by modifying a method described previously (8), using a DNA scaffold containing the −10 and consensus discriminator elements (23) in the nontemplate strand (Fig. 1*A*). To obtain the crystals of the *de novo* transcription initiation complex, the holoenzyme-promoter DNA crystals were soaked with the first two iNTPs that base pair with the +1 and +2 bases of template DNA. To maintain the transcription bubble during crystallization of the holoenzyme-promoter DNA complex, previous studies used a noncomplementary template strand sequence upstream of the transcription start site (8, 9). Our previous structural analysis of the N4 phage RNAP *de novo* transcription complex (7) identified that the template DNA nucleotide at the −1 position participates in binding the first iNTP by a base stacking interaction. Therefore, in this study, we changed the template DNA sequence at the −1 position to a guanine base (the template-strand sequence 3′-GTGA-5′; the transcription start site is underlined), thus mimicking the natural sequence of most bacterial promoters recognized by the primary σ factor (24). By incubating the crystals with ATP and a nonhydrolyzable CTP analog, CMPCPP (cytidine-5′-[(α,β)–methyleno]triphosphate), we obtained the *de novo* transcription initiation complex structure with the two iNTPs bound at the active site, but without phosphodiester bond formation. The structure was solved by molecular replacement at 2.9 Å resolution. The structure showed strong unbiased $F_o - F_c$ electron densities corresponding to the first and second iNTPs (Fig. 2*A*, supplemental Movie S1). The overall structure of RNAP was identical with that of the search model (Protein Data Bank code 4G7H) (8). However, important differences included conformational changes in the trigger loop of the β' subunit (residues 1233–1255) and the amino acid side chains of both the β and β' subunits around the active site (Figs. 2, *B* and *C*).

The First iNTP Binding Site in the de Novo Transcription Initiation Complex—Binding of the first and second iNTPs involves interactions with residues around the active site that are unique to *de novo* transcription and other interactions that are common to RNA elongation. The first iNTP occupies the *i* site in the *de novo* transcription initiation complex, which overlaps with the RNA 3′-end binding site in the elongation complex (Fig. 2*C*) (1). The *de novo* transcription-specific interactions between RNAP and iNTPs appear to concentrate on the triphosphate of the first iNTP, which is ATP in our crystal structure, with multiple salt bridges with the β subunit residues, including 1) Gln-β567 and His-β999 (interact with a nonbridging oxygen of γ-phosphate; correspond to *E. coli* RNAP residues Gln-β688 and His-β1237); and 2) Lys-β838 (interacts with a nonbridging oxygen of α-phosphate; corresponds to

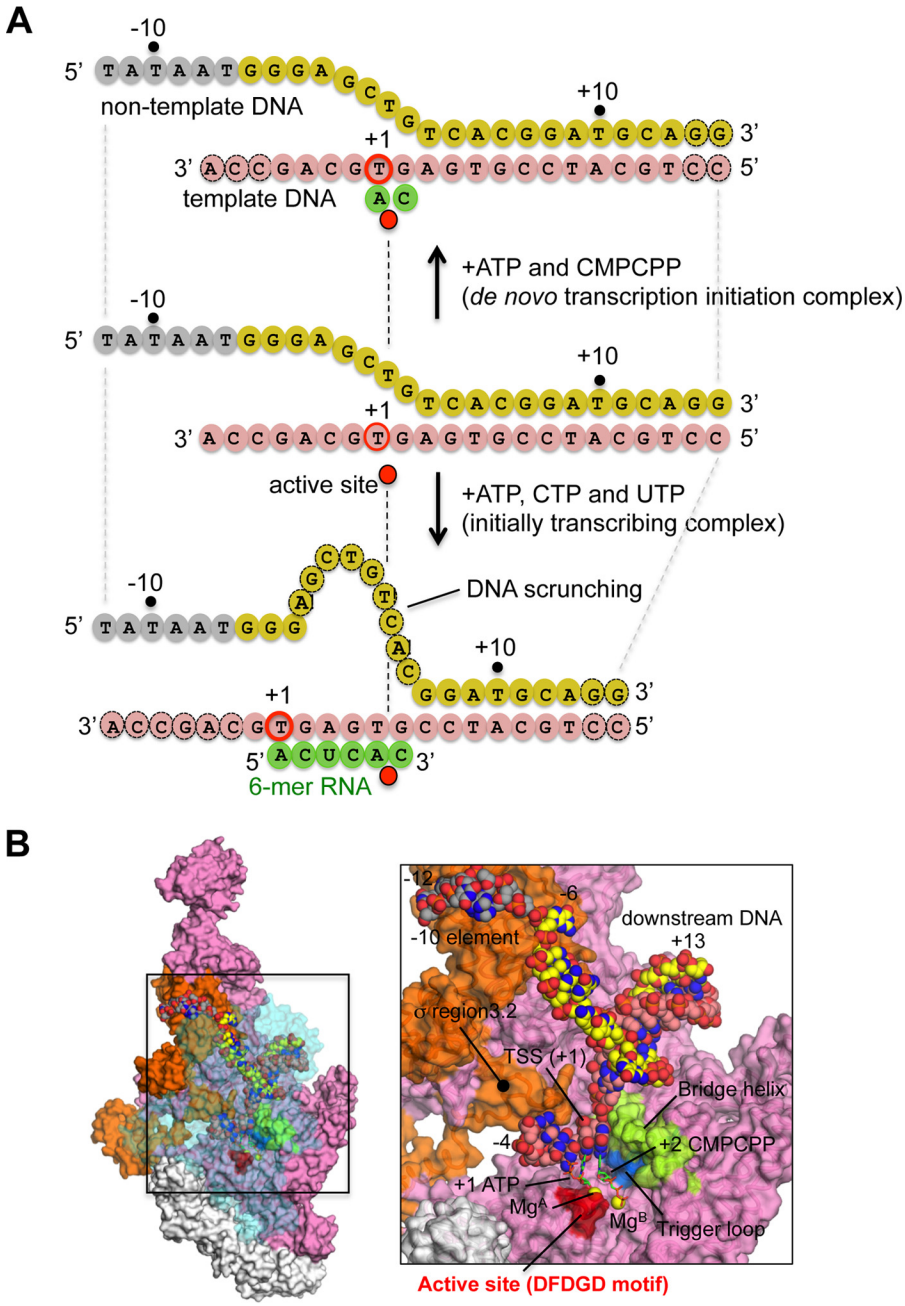
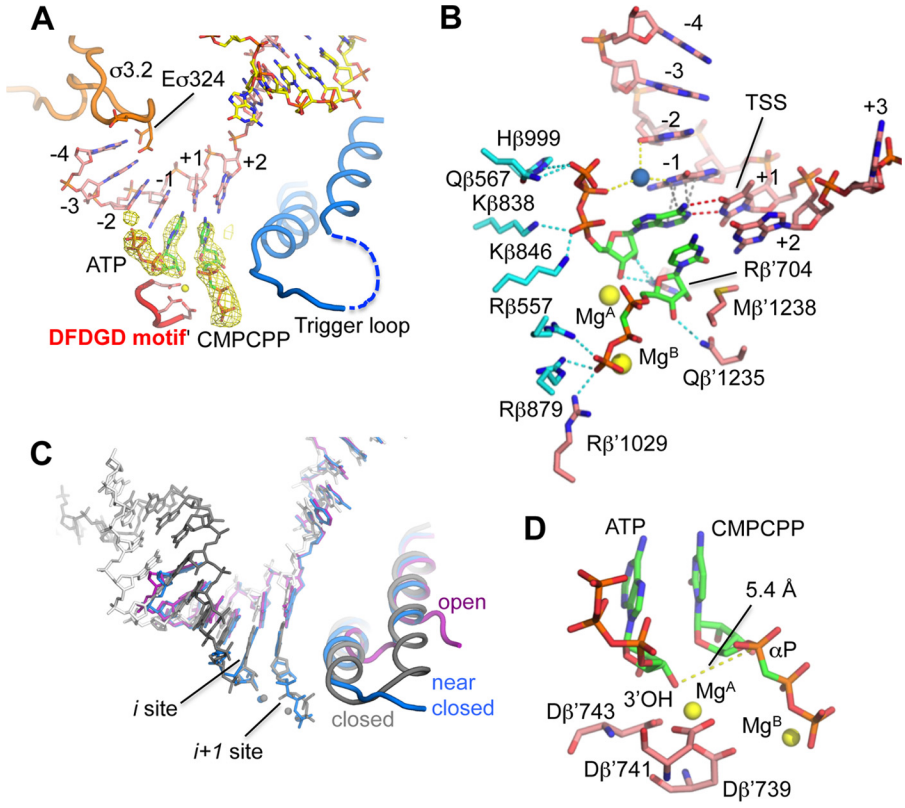


FIGURE 1. **Structure of the transcription initiation complex.** *A*, schematic of the *de novo* transcription initiation complex (*top*) and the initially transcribing complex (*bottom*) obtained from the RNAP-promoter complex. Sequence of the nucleic acid scaffold used for holoenzyme-promoter DNA complex crystallization is shown in the *middle* (*pink*, template DNA; *yellow*, non-template DNA; *gray*, −10 element). NTPs and 6-mer RNA in the transcription complexes are depicted by *green circles*. Positions of the RNAP active site (*red sphere*) as well as template DNA transcription start site (TSS, +1, *red circle*) are indicated. DNA bases indicated by *black dashed circles* are disordered in the crystal structures. In a standard format describing RNAP promoter sequence, the transcription start site, which base pairs with the first iNTP and encodes the 5′-end of the RNA chain, is designated +1. DNA positions downstream from the transcription start site increase as positive numbers from +1 and DNA positions upstream of the start site increase as negative numbers counting from −1. *B*, overall structure of the *de novo* transcription initiation complex. *T. thermophilus* RNAP is depicted as a molecular surface model. Each subunit of RNAP is depicted with a unique color (*white*, α; *cyan*, β; *pink*, β′; *orange*, σ). DNA is depicted as a sphere model with the same color scheme as described in *A*. *Right panel* shows a magnified view of the *boxed* region in the *left panel*. For clarity, the β subunit was made transparent in the left panel and removed in the right panel. Several key motifs discussed in the text are highlighted (see also supplemental Movie S1).

E. coli residue Lys-β1065) (Fig. 2, *B* and *E*). Electron density map for the Lys-β846 side chain (*E. coli* residue Lys-β1073) is not well defined (data not shown), but it may be positioned near a nonbridging oxygen of α-phosphate for making an additional salt bridge. Arg-704 of the β′ subunit (Arg-β′704; *E. coli* residue Arg-β′425) forms salt bridges with 2′-OH and 3′-OH of the ATP and may therefore participate in discrimination between NTP and dNTP. All residues interacting with the ATP triphosphate are absolutely conserved in all cellular RNAPs (Fig. 2*E*), suggesting a universal mechanism for accommodating the first iNTP at the enzyme active site to establish *de novo* transcription.

The structure is consistent with previous biochemical studies showing that residues Lys-β1065, Lys-β1073, and His-



E

Bacteria[a]		Eukaryote[b]			Archaea[c]	
Tth	*E. coli*	Pol I	Pol II	Pol III	*Tko*[d]	*Sso*[e]
Q567	Q688	Q724	Q776	Q708	Q677	Q675
K838	K1065	K916	K979	K911	K881	K877
K846	K1073	K924	K987	K919	K889	K885
H999	H1237	H1038	H1097	H1029	H999	H995

[a]β subunit of bacterial RNAP
[b]Second largest subunits of *Saccharomyces cerevisiae* RNAPs (A135 in Pol I, Rbp2 in Pol II, C128 in Pol III)
[c]Second largest subunits of archaeal RNAPs (RpoB)
[d]*Thermococcus kodakarensis*
[e]*Sulfolobus sulfataricus*

FIGURE 2. **First and second iNTP binding sites of the *de novo* transcription initiation complex.** *A*, active site structure of the *de novo* transcription initiation complex. DNA, iNTPs, and amino acid side chains are shown as stick models, and the trigger loop, DFDGD motif, and the σ region 3.2 (σ3.2) are shown as worm models and labeled. The disordered region of the trigger loop is shown as a *dashed line*. $F_o - F_c$ electron density maps (*yellow mesh*) showing ATP (first iNTP) and CMPCPP (second iNTP) are superposed on the final model. *B*, the first iNTP (ATP) binds at the active site through multiple interactions, including base pairing with the +1 DNA base (*red dashed lines*), base stacking with the −1 purine base (*gray dashed lines*), water-mediated interactions (*yellow dashed lines*), and salt bridges with side chains (*cyan dashed lines*). Amino acid side chains involved in the second iNTP (CMPCPP) binding are also indicated. The transcription start site (*TSS*) is indicated. *C*, comparison of the holoenzyme-promoter DNA (Protein Data Bank code 4G7H, *magenta*), *de novo* transcription (*blue*), and transcription elongation complexes (Protein Data Bank code 2O5J, *gray* and *white*). RNAP structures were superposed at their active site domains of the β' subunit. The *i* and *i*+1 sites are indicated. Trigger loops (β' residues 1222–1265) of these structures are shown as tube models. *D*, ATP and CMPCPP bound in the active site. The aspartate residues of the DFDGD motif of the β' subunit coordinating the Mg^{2+} (*yellow spheres*) are labeled. The distance between the 3′-OH of ATP and the α-phosphate of CMPCPP is 5.4 Å as indicated. *E*, amino acid residues involved in the first iNTP binding are conserved in all cellular RNAPs.

β1237 of *E. coli* RNAP (counterparts of *T. thermophilus* residues Lys-β838, Lys-β846, and His-β999, respectively) are close to the binding site of the first iNTP and that their substitutions impaired transcription (25, 26). To further analyze the functions of the amino acid residues interacting with the first iNTP, we made four *E. coli* RNAP mutants, including Lys-β1065A and Lys-β1065A/Lys-β1073A, for residues interacting with α-phosphate, and His-β1237A and His-β1237A/Gln-β688A, for residues interacting with γ-phosphate, and tested their transcription activities on the T7A1 promoter (Fig. 3). The mutations resulted in major defects in transcription initiation in the presence of low concentrations of NTPs, resulting in decreased abortive and full-length RNA synthesis (Fig. 3*B*, *lanes 1–6*). The activities of the Lys-β1065A and His-β1237A RNAPs were partially recovered by adding an initiating dinucleotide primer (*lanes 7–12*) and were further rescued in the presence of high concentrations of NTPs (*lanes 13–18* and *19–24*), suggesting that the mutations primarily affect the initiation step of transcription. Importantly, the mutations also affected the pattern of abortive products synthesized during

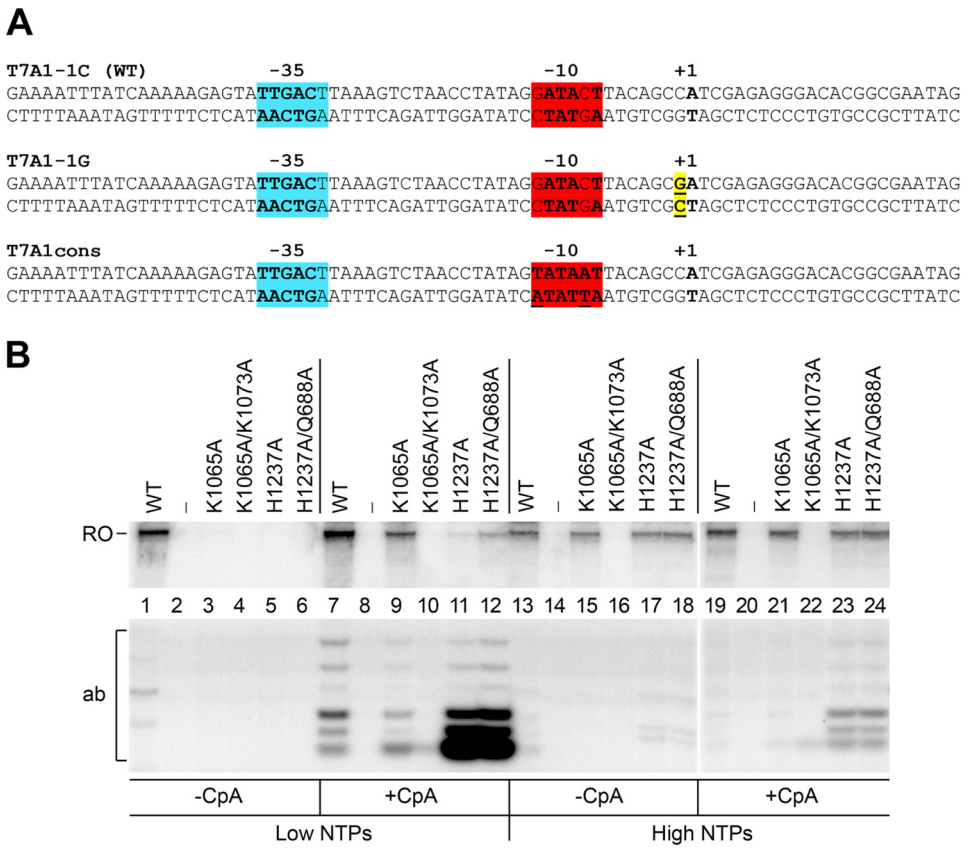
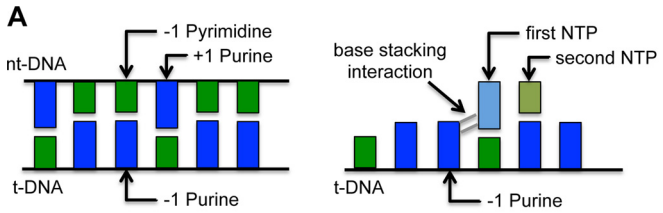


FIGURE 3. **Transcription activities of wild-type and mutant RNAPs on the T7A1 promoter.** *A*, the sequences of the wild-type T7A1 promoter and its variants used in this study. Positions of the −35, −10 promoter elements, and the transcription start site (+1) are indicated. The T7A1 promoter variant has a base substitution at the −1 position (highlighted in *yellow*); the T7A1 consensus (T7A1cons) promoter contains the consensus −10 element (TATAAT). *B*, transcription was performed in the presence of low (10 μM all four NTPs with addition of α-[^{32}P]UTP) or high (100 μM ATP, GTP, CTP, and 10 μM UTP) concentrations of NTPs with or without initiating primer CpA (25 μM), corresponding to positions −1/+1 of the promoter. The positions of the run-off (*RO*) and abortive (*ab*) RNA products are indicated.

initiation (Fig. 3*B*; see below). The double mutant His-β1237A/Gln-β688A did not differ significantly from the His-β1237A RNAP; however, the double mutant Lys-β1065A/Lys-β1073A was essentially inactive at all conditions tested, indicating that residue Lys-β1073A is essential for initiation. The results highlight the relevance of these residues during *de novo* transcription initiation and are consistent with their proposed roles in the first iNTP binding.

Based on the previous observations on the role of σ region 3.2 in iNTP binding (3, 27), we next inspected possible interactions between this region and the iNTPs loaded in the active site. The closest distance between the σ region 3.2 (Glu-324 side chain) and the γ-phosphate of the *i* site bound ATP is ~10.5 Å; at this distance, the σ region 3.2 is positioned near the upstream template DNA bases −3/−4 (Fig. 2*A*). This rules out any direct interaction between this region and the first iNTP during *de novo* initiation and suggests that its stimulatory effect on iNTP binding is indirect.

From the structure, we noticed a base stacking interaction between the guanine template position −1 and the *i* site bound ATP (supplemental Movie S1); also the N-1 of the guanine base and the β-phosphate of ATP are connected by a water molecule (Fig. 2*B*). The base stacking interaction would be minimized in the case of template DNA containing a pyrimidine base at the −1 position (Fig. 4*A* and supplemental Movie S1). These additional interactions may contribute to stable binding of the ATP during *de novo* transcription. To test this hypothesis, we measured the effects of a G-to-C substitution at the −1 template position on the K_m of iNTPs on the T7A1 promoter (the wild-type template strand sequence 3′-GTAG-5′, the transcription start site is underlined, Fig. 3*A*). We found that this substitution greatly increased apparent K_m values for both +1ATP and +2UTP. In fact, the effect for the +2UTP was even stronger than for the ATP (78.8- and 8.9-fold changes compared with the wild-type T7A1 promoter, respectively) (Fig. 4*B*). Thus, the interaction between the −1 template base and the +1ATP may not only contribute to the +1ATP binding, but also stabilize the template DNA strand and thus stimulate the +2UTP binding. To further test this hypothesis, we also measured K_ms for *E. coli* RNAP with a σ region 3.2 deletion (Δ513–519), which decreases the iNTP binding due to destabilizing the template DNA strand (20). Deletion of σ region 3.2 increased K_ms for both +1ATP and +2UTP on the wild-type T7A1 promoter. However, the promoter substitution only slightly increased K_ms for the mutant RNAP (3- and 2.5-fold for +1ATP and +2UTP, respectively), suggesting that both the σ region 3.2 deletion and the −1 nucleotide substitution disturb the template strand conformation.



B

ATP + UTP → pppApU			
K_M		T7A1 (T7A1-1C)	T7A1 variant (T7A1-1G)
ATP	σ^{70}	161 ± 48.9	1,440 ± 260
		1	8.9
	σ^{70} Δ513-519	800 ± 175	2,540 ± 280
		1	3.2
UTP	σ^{70}	10.4 ± 2.3	820 ± 70
		1	78.8
	σ^{70} Δ513-519	570 ± 140	1,440 ± 115
		1	2.5

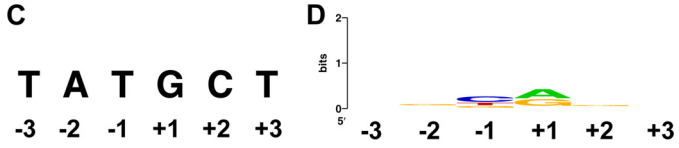


FIGURE 4. **Role of the base stacking interaction between the first iNTP and template DNA in transcription initiation.** *A*, schematic representation of the base stacking interaction between the −1 template DNA purine base and the first iNTP purine base. (see also supplemental Movie S1). *B*, apparent K_m values for the initiating substrates on the wild-type T7A1 promoter (−1C nontemplate sequence) and its variant (−1G non-template sequence) for RNAPs containing wild-type or Δ513–519 σ^{70} factors. The numbers in *blue* and *boldface* type show changes in K_m values for the T7A1 promoter variant relative to the wild-type sequence. *C*, non-template DNA sequence around the transcription start sites of human Pol I, which is adapted from Ref. 35. The transcription start site is +1. *D*, sequence conservation of the non-template DNA around transcription start site of human Pol II. The sequence logo was prepared using experimentally determined non-redundant collection of human Pol II promoters, for which the transcription start site (+1) has been determined experimentally. The majority of human Pol II promoters contain −1 pyrimidine and +1 purine bases in the non-template DNA.

Second iNTP Binding Site in the de Novo Transcription Initiation Complex—The second iNTP, CMPCPP in our crystal structure, is positioned through base pairing with template DNA and interacts with the β′ subunit trigger loop and basic residues (Arg-β557, Arg-β879, and Arg-β′1029) on the rim of the secondary channel (Fig. 2*B*). In the holoenzyme-promoter DNA complex, the trigger loop is in an open conformation and the tip of the loop (residues 1238–1251) is disordered (8), whereas in the elongation complex with an NTP substrate at the *i*+1 site, the trigger loop is in the closed conformation and forms two trigger helixes without disordered regions (Fig. 2*C*) (1). In the *de novo* transcription initiation complex, even though the trigger loop is in a more closed conformation compared with the holoenzyme-promoter DNA complex, the middle portion of the trigger loop (residues 1246–1251) is still disordered. Met-β′1238 and Gln-β′1235 residues from the N-terminal α-helix of the trigger loop reach out to interact with the base and 3′-OH of CMPCPP, respectively. However, because the trigger helix is not fully folded, there is no direct interaction of the triphosphate group of the second iNTP with Arg-β′1239 and His-β′1242, causing it to attain a preinsertion conformation (Fig. 2*D*). In particular, the α-phosphate of CMPCPP is ~5.4 Å away from the 3′-OH of ATP forming a nonreactive state. The catalytic metal Mg^A is bound stably at the active site through the aspartate triad (residues at 739, 741, and 743) of the DFDGD motif of the β′ subunit. However, Mg^B is too far away for being coordinated by the aspartate triad (Fig. 2*D*) and is primarily coordinated by the β- and γ-phosphates of CMPCPP. This nucleotide position is similar to the preinsertion state observed in crystal structures of the *T. thermophilus* elongation complex with the inhibitor streptolydigin (1), the eukaryotic Pol II elongation complex containing GMPCPP (guanine-5′–[(α,β)–methyleno]triphosphate) (28), and the Pol II initially transcribing complexes containing short RNA (2~7 nt in length) and AMPCPP (adenine-5′–[(α,β)–methyleno]triphosphate) (29). Similar to these structures, using the non-hydrolyzable CMPCPP for preparing the bacterial *de novo* transcription initiation complex in this study might affect the coordination of the iNTP and metals at the active site, thus positioning the *i*+1 nucleotide in the precatalytic conformation, which likely corresponds to a natural intermediate in formation of the catalytically competent transcription inititiation complex. During *de novo* transcription initiation in the presence of natural iNTP substrates, the trigger loop would then adopt the completely closed conformation, forming the trigger helixes to push the iNTP at the *i*+1 site closer to the 3′-OH at the *i* site and resulting in the formation of the first phosphodiester bond.

Structure of the Initially Transcribing Complex of RNAP Holoenzyme Containing 6-Mer RNA—Crystal structures of transcription complex containing short RNAs of 2 to 7 nt in length have been determined for the Pol II system. However, these complexes were prepared by incubating Pol II or Pol II/TFIIB with DNA template and synthetic RNA oligonucleotides (29–31). Therefore, these complexes lack the triphosphate group at the 5′-end of RNA that may play important roles in the early stages of transcription such as DNA/RNA hybrid stabilization, separation of RNA from template DNA after reaching the full hybrid length and removal of the transcription initiation factor from RNAP. To obtain the structure of an early-stage transcription complex containing the σ factor and the natural form of RNA containing the 5′-triphosphate group, we prepared the initially transcribing complex by *in crystallo* approach. For this purpose, we soaked the holoenzyme-promoter DNA complex crystals with ATP, CTP, and UTP, resulting in the synthesis of a 6-mer RNA (the template sequence is 3′-GTGAGTGC-5′, the transcription start site is underlined) (Fig. 1*A*). The structure was solved at 3 Å resolution, and it showed a continuous electron density for the 6-mer RNA synthesized along the template DNA (Fig. 5, *A* and *B*, and supplemental Movie S1), indicating that *T. thermophilus* RNAP was active in the crystalline state and capable of transcribing RNA to the expected length.

The RNA transcript remains in the pretranslocated state, with its 3′-end bound in the *i*+1 site; however, the pyrophosphate by-product is not visible, and the middle region of the trigger loop is disordered (residues 1239–1253) (Fig. 5*B*). Further movement of the DNA/RNA hybrid to a post-translocated state may be constrained due to steric hindrance between the

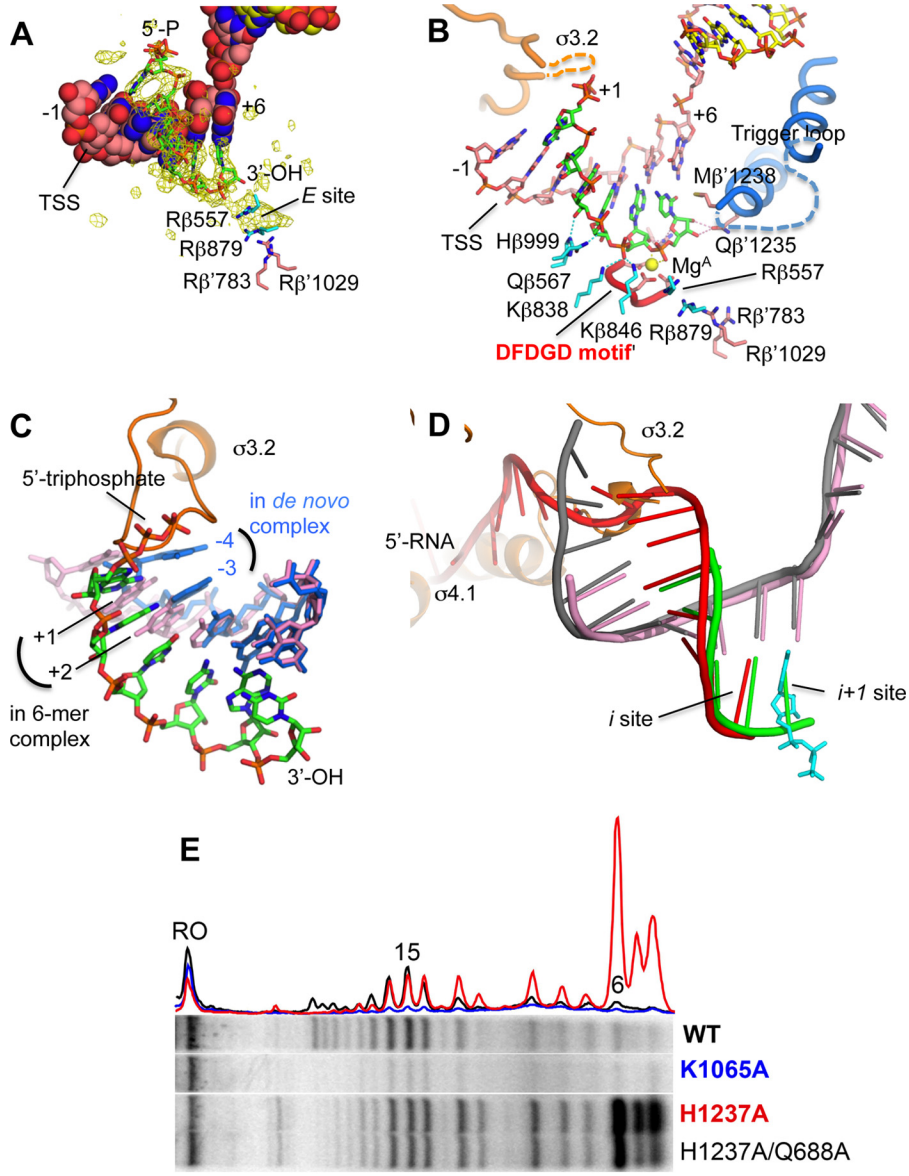


FIGURE 5. **The initially transcribing complex of RNAP holoenzyme containing 6-mer RNA.** *A*, $F_o - F_c$ electron density map (*yellow mesh*) showing 6-mer RNA and a NTP bound at the *E* site of the initially transcribing complex. The final model of 6-mer RNA and amino acid side chains involved in NTP binding at the *E* site are shown as stick models and labeled. The DNA is depicted as CPK model. *B*, active site structure. DNA, RNA, and amino acid side chains contacting the 6-mer RNA are shown as stick models. Disordered regions of the trigger loop (residues 1239–1253) and the σ region 3.2 (σ3.2, residues 321–327) are indicated by *dashed lines*. The transcription start site (*TSS*) is indicated. *C*, comparison of the *de novo* transcription (*blue*, template DNA; *orange*, σ region 3.2) and the 6-mer RNA initially transcribing complexes (*pink*, template DNA; *green*, 6-mer RNA). RNAP structures were superposed on the active site domains of the β′ subunit. *D*, model of an extended RNA transcript placed in the RNA exit channel explains σ release. Structures of the initially transcribing (*pink*, template DNA; *green*, 6-mer RNA; *orange*, σ factor) and the transcription elongation complexes (Protein Data Bank code 2O5J, *gray*, template DNA; *red*, RNA; *cyan*, incoming NTP) were superposed on the active site domains of the β′ subunit. The *i* and *i*+1 sites are indicated. *E*, effects of the *E. coli* RNAP mutations on abortive synthesis. Transcription was performed on the T7A1cons promoter at high NTP concentrations in the presence of the CpA primer. Positions of the run-off (*RO*) and abortive products are indicated. The profiles *above* the gel show distribution of the RNA products for the wild-type, K1065A, and H1237A RNAPs (*black*, *blue*, and *red*, respectively).

5′-end of RNA and the σ region 3.2 (see below), and the ejection of σ region 3.2 may be restricted due to crystal packing. A bigger lobe of electron density appears next to the RNA 3′-end (Fig. 5*A* and supplemental Movie S1). This density is close to the rim of the secondary channel and overlaps with the temporary NTP binding *E* site defined earlier in Pol II (32), which is distinct from the preinsertion site (1).

The downstream duplex DNA in the initially transcribing complex is shifted 6 base pairs upstream due to RNA transcription, whereas the contact between the upstream −10 DNA element and the σ region 2 is maintained. Thus, initial RNA transcription renders nontemplate DNA bases from −3 to +6 disordered due to scrunching of the nontemplate strand (Fig. 1*A* and supplemental Movie S1).

Compared with the *de novo* transcription initiation complex, the initially transcribing complex with 6-mer RNA also reveals changes in the σ region 3.2 and the template DNA near the σ region 3.2 (Fig. 5*C*). In the *de novo* complex, the template DNA bases at the −4/−3 positions are flipped out to make contacts with the acidic residues of the σ region 3.2, whereas in the

initially transcribing complex, DNA bases at corresponding positions upstream of the active site base pair with the nascent 6-mer RNA. The 5′-end of the RNA transcript reaches σ region 3.2 causing the tip of this region (residues 321–327) to become disordered (Fig. 5*B*). This is likely due to the charge repulsion between the acidic cluster of the σ region 3.2 and the 5′-triphosphate of RNA. When RNAP continues transcription, the σ region 3.2 has to be pushed further to accommodate longer RNA transcript extending toward the σ region 4.1, ultimately resulting in σ ejection from the RNA exit channel (Fig. 5*D*).

Overall, the DNA/RNA hybrid in this initially transcribing complex is a typical A-form duplex with no intrinsic tilting that was observed earlier in the Pol II initially transcribing complexes (Fig. 5*D*) (29, 31). As the DNA/RNA hybrid extends in length, the initial contacts between RNAP and the triphosphate group of the first iNTP are lost. The positively charged and polar residues involved in interactions with the triphosphate group of the first iNTP in the *de novo* transcription initiation complex (Gln-β567, Lys-β838, Lys-β846, and His-β999) are now involved in interactions with internal phosphate and ribose groups at positions −3/−4 of the nascent RNA, suggesting their role in stabilization of the short RNA-DNA hybrid (Fig. 5*B*). To test the role of these residues in the extension of short RNAs, we analyzed abortive transcription by the wild-type and mutant *E. coli* RNAP variants on a consensus T7A1cons promoter (Fig. 3*A*), which is characterized by a very high efficiency of abortive synthesis due to strong RNAP-promoter interactions (20). Transcription was performed at high NTP concentrations and in the presence of the dinucleotide primer, to ensure efficient *de novo* initiation by the mutant RNAPs. Surprisingly, substitution Lys-β1065A decreased abortive RNA synthesis (Fig. 5*E*), suggesting that this residue may directly or indirectly disturb short RNA extension in the wild-type RNAP, while playing a role in the first bond formation (see above). In contrast, substitutions His-β1237A and His-β1237A/Gln-β688A greatly stimulated the synthesis of ≤10-nt abortive RNAs, but did not affect longer abortive products. Similarly, these mutations increased the amounts of short abortive RNAs synthesized on the wild-type T7A1 promoter (Fig. 3*B*, *lanes 11* and *12* and lanes *23* and *24*). Therefore, residue His-β1237 plays an important role in stabilization of short nascent RNAs in the RNAP active site, until the full DNA/RNA hybrid is formed and the RNA reaches the RNA exit channel.

DISCUSSION

In this study, we report the structures of bacterial transcription initiation complexes with iNTPs or short RNA product bound in the active site of RNAP. In comparison with existing structures, they reveal essential interactions of the iNTPs with template DNA and for the first time highlight initial steps of RNA extension by bacterial RNAP. Furthermore, we provide biochemical support to our structural findings and demonstrate the importance of the observed contacts of iNTPs and RNA with RNAP and the DNA template for transcription initiation.

The binding of the first iNTP is a special feature of all cellular RNAPs that are capable of primer-independent *de novo* transcription initiation. A network of salt-bridge and hydrogen bonds between the triphosphate of the first iNTP and RNAP side chains is critical for its binding. At the same time, the positions of the ribose and the base of the first iNTP are identical to the *i* site in the elongation complex. Furthermore, the position of the second iNTP in the promoter complex corresponds to the position of the incoming NTP bound in preinsertion state in the elongation complex, indicating that the mechanisms of the phosphodiester bond formation are identical in the transcription initiation and elongation. Similar positions of the iNTPs were observed in the *T. thermophilus* RNAP *de novo* transcription initiation complex structure, which has been reported by Ebright and co-workers (9) as a part of the analysis of the GE23077 antibiotic targeting bacterial RNAP while our manuscript was in preparation. We now provide strong biochemical evidence in support of the importance of the observed iNTP contacts with RNAP and DNA for *de novo* transcription initiation.

The structures we obtained by using template DNA containing a purine nucleotide at the −1 position revealed a stacking interaction between the −1 purine base in template DNA and the first purine iNTP. These stacking interactions were not observed in previously reported structures because they used a suboptimal template sequence with a template pyrimidine at −1 position. We showed that the base stacking interaction plays major roles in the binding of not only the first iNTP but also the second iNTP, likely as a result of template stabilization (Fig. 4*B*). This purine-purine base stacking during *de novo* transcription initiation explains the preference of the purine nucleotide at the transcription start site and pyrimidine nucleotide at the preceding position in the non-template strand in bacterial promoters (Fig. 4*A*) (24, 33, 34). For example, in the *Mycobacterium tuberculosis* genome, the occurrences of pyrimidines and purines at the −1 and +1 non-template DNA bases are >70 and 80%, respectively (33). It remains to be established whether the presence of suboptimal −1 nucleotides in the minority of promoters has any specific role in transcription regulation in bacteria. The preferable pyrimidine (−1) and purine (+1) combination is also found in the majority of eukaryotic Pol I (35) and Pol II promoters (Fig. 4, *C* and *D*) (36). A similar base pair stacking interaction was noticed in the structure of the N4 phage RNAP *de novo* transcription initiation complex (7), indicating that this mechanism may be universal.

Although the role of the σ factor in promoter recognition is well established, the structural basis for its role in the binding of iNTPs remained unknown. Our structure reveals the absence of any direct contacts between the σ region 3.2 and the first iNTP. Rather, this region seems to guide the path of template DNA, thereby positioning the DNA bases for ideal binding with the iNTPs. Therefore, deletion of the σ region 3.2 likely disturbs template DNA binding, especially around the transcription start site, allowing it to adopt a floppy conformation that hampers stable binding of the iNTPs (3, 20, 37). The eukaryotic Pol II counterpart of the σ region 3.2 is the B-reader of TFIIB, which analogously inserts its α-helix deep into the DNA binding cleft close to the active site of Pol II to ensure proper template DNA

positioning as well as participate in the transcription start site selection (30, 38, 39).

When RNAP transcribes a 6-mer RNA, the RNA 5′-end reaches the tip of the σ region 3.2, causing it to become disordered, most likely due to repulsion between the RNA 5′-triphosphate and the acidic cluster at the σ region 3.2 tip. The 5′-triphosphate group is also partially disordered, indicating that there are no stable interactions between the σ region 3.2 and RNA. Our initially transcribing complex structure indicates that synthesis of a short RNA transcript is the first step of the σ ejection process that is likely followed by ejection of σ region 4 from the RNA exit channel and allows for the initiation-to-elongation transition. Accordingly, deletion of the σ factor from its region 3.2 to C terminus eliminates abortive RNAs (40), whereas deletion of the σ region 3.2 reduces only 5∼9-mer abortive RNAs (20). Similarly, the B-reader of factor TFIIB in eukaryotic Pol II should be extruded from the RNA channel after RNA reaches 6 nucleotides in length, revealing striking similarities to the bacterial system (23).

The initially transcribing complexes of RNAP holoenzyme containing short RNAs, usually 2–12 nucleotides in length, are unstable in solution resulting in abortive initiation (10, 11). In this study, we demonstrated that the *T. thermophilus* RNAP is active in crystallized state and is capable of synthesizing up to 6-mer RNA. In the crystallized state, RNAP motions are restricted by the crystal packing, likely stabilizing the initially transcribing complex and allowing for preparation of highly homogenous complexes. The *in crystallo* transcription system can be used to study RNA transcription by Raman crystallography and time-resolved trigger-freeze crystallography. Analogous to the studies performed on single-subunit RNAPs (41, 42), these new experimental approaches may allow to trace events in not only initial transcription but also during the NTP addition cycle in cellular RNAPs.

Acknowledgments—We thank Drs. Irina Artsimovitch and Shun-ichi Sekine for providing information of culturing T. thermophilus cells and RNAP purification and for plasmids. We thank the staff at Penn State Fermentation Facility for supporting T. thermophilus cell culture and the MacCHESS for supporting crystallographic data collection. We thank Dr. Lilian Hsu for critical reading of the manuscript. Figs. were prepared using PyMOL.

REFERENCES

1. Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007) Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448,** 163–168
2. McClure, W. R., Cech, C. L., and Johnston, D. E. (1978) A steady state assay for the RNA polymerase initiation reaction. *J. Biol. Chem.* **253,** 8941–8948
3. Kulbachinskiy, A., and Mustaev, A. (2006) Region 3.2 of the sigma subunit contributes to the binding of the 3′-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J. Biol. Chem.* **281,** 18273–18276
4. Gaal, T., Bartlett, M. S., Ross, W., Turnbough, C. L., Jr., and Gourse, R. L. (1997) Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science* **278,** 2092–2097
5. Turnbough, C. L., Jr., and Switzer, R. L. (2008) Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. *Microbiol. Mol. Biol. Rev.* **72,** 266–300
6. Kennedy, W. P., Momand, J. R., and Yin, Y. W. (2007) Mechanism for *de novo* RNA synthesis and initiating nucleotide specificity by T7 RNA polymerase. *J. Mol. Biol.* **370,** 256–268
7. Gleghorn, M. L., Davydova, E. K., Basu, R., Rothman-Denes, L. B., and Murakami, K. S. (2011) X-ray crystal structures elucidate the nucleotidyl transfer reaction of transcript initiation using two nucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 3566–3571
8. Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M. X., Arnold, E., and Ebright, R. H. (2012) Structural basis of transcription initiation. *Science* **338,** 1076–1080
9. Zhang, Y., Degen, D., Ho, M. X., Sineva, E., Ebright, K. Y., Ebright, Y. W., Mekler, V., Vahedian-Movahed, H., Feng, Y., Yin, R., Tuske, S., Irschik, H., Jansen, R., Maffioli, S., Donadio, S., Arnold, E., and Ebright, R. H. (2014) GE23077 binds to the RNA polymerase 'i' and 'i+1' sites and prevents the binding of initiating nucleotides. *eLife* **3,** e02450
10. Hsu, L. M. (2002) Promoter clearance and escape in prokaryotes. *Biochim. Biophys. Acta* **1577,** 191–207
11. Hsu, L. M. (2009) Monitoring abortive initiation. *Methods* **47,** 25–36
12. Murakami, K. S., and Darst, S. A. (2003) Bacterial RNA polymerases: the wholo story. *Curr. Opin. Struct. Biol.* **13,** 31–39
13. Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S., and Ebright, R. H. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314,** 1144–1147
14. Revyakin, A., Liu, C., Ebright, R. H., and Strick, T. R. (2006) Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314,** 1139–1143
15. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326
16. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674
17. Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P., and Adams, P. D. (2010) Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr. D* **66,** 1153–1163
18. Murakami, K. S. (2013) X-ray crystal structure of *Escherichia coli* RNA polymerase sigma70 holoenzyme. *J. Biol. Chem.* **288,** 9126–9134
19. Pupov, D., Miropolskaya, N., Sevostyanova, A., Bass, I., Artsimovitch, I., and Kulbachinskiy, A. (2010) Multiple roles of the RNA polymerase β′ SW2 region in transcription initiation, promoter escape, and RNA elongation. *Nucleic Acids Res.* **38,** 5784–5796
20. Pupov, D., Kuzin, I., Bass, I., and Kulbachinskiy, A. (2014) Distinct functions of the RNA polymerase sigma subunit region 3.2 in RNA priming and promoter escape. *Nucleic Acids Res.* **42,** 4494–4504
21. Dreos, R., Ambrosini, G., Cavin Périer, R., and Bucher, P. (2013) EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era. *Nucleic Acids Res.* **41,** D157–164
22. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190
23. Feklistov, A., Barinova, N., Sevostyanova, A., Heyduk, E., Bass, I., Vvedenskaya, I., Kuznedelov, K., Merkiene, E., Stavrovskaya, E., Klimasauskas, S., Nikiforov, V., Heyduk, T., Severinov, K., and Kulbachinskiy, A. (2006) A basal promoter element recognized by free RNA polymerase σ subunit determines promoter recognition by RNA polymerase holoenzyme. *Mol. Cell* **23,** 97–107
24. Shultzaberger, R. K., Chen, Z., Lewis, K. A., and Schneider, T. D. (2007) Anatomy of *Escherichia coli* σ^{70} promoters. *Nucleic Acids Res.* **35,** 771–788
25. Sagitov, V., Nikiforov, V., and Goldfarb, A. (1993) Dominant lethal mutations near the 5′ substrate binding site affect RNA polymerase propagation. *J. Biol. Chem.* **268,** 2195–2202
26. Mustaev, A., Kashlev, M., Lee, J. Y., 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
27. Severinov, K., Fenyö, D., Severinova, E., Mustaev, A., Chait, B. T., Goldfarb, A., and Darst, S. A. (1994) The sigma subunit conserved region 3 is part of "5′-face" of active center of *Escherichia coli* RNA polymerase. *J. Biol. Chem.* **269,** 20826–20828
28. Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D., and Kornberg,

R. D. (2006) Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127,** 941–954
29. Cheung, A. C., Sainsbury, S., and Cramer, P. (2011) Structural basis of initial RNA polymerase II transcription. *EMBO J.* **30,** 4755–4763
30. Sainsbury, S., Niesser, J., and Cramer, P. (2013) Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* **493,** 437–440
31. Liu, X., Bushnell, D. A., Silva, D. A., Huang, X., and Kornberg, R. D. (2011) Initiation complex structure and promoter proofreading. *Science* **333,** 633–637
32. Westover, K. D., Bushnell, D. A., and Kornberg, R. D. (2004) Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119,** 481–489
33. Cortes, T., Schubert, O. T., Rose, G., Arnvig, K. B., Comas, I., Aebersold, R., and Young, D. B. (2013) Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell Rep.* **5,** 1121–1131
34. Rhodius, V. A., and Mutalik, V. K. (2010) Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, σ^E. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 2854–2859
35. Massin, P., Rodrigues, P., Marasescu, M., van der Werf, S., and Naffakh, N. (2005) Cloning of the chicken RNA polymerase I promoter and use for reverse genetics of influenza A viruses in avian cells. *J. Virol.* **79,** 13811–13816
36. Butler, J. E., and Kadonaga, J. T. (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* **16,** 2583–2592
37. Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L., and Darst, S. A. (2002) Structure of the bacterial RNA polymerase promoter specificity σ subunit. *Mol. Cell* **9,** 527–539
38. Liu, X., Bushnell, D. A., Wang, D., Calero, G., and Kornberg, R. D. (2010) Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* **327,** 206–209
39. Kostrewa, D., Zeller, M. E., Armache, K. J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009) RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462,** 323–330
40. Murakami, K. S., Masuda, S., and Darst, S. A. (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296,** 1280–1284
41. Basu, R. S., and Murakami, K. S. (2013) Watching the bacteriophage N4 RNA polymerase transcription by time-dependent soak-trigger-freeze X-ray crystallography. *J. Biol. Chem.* **288,** 3305–3311
42. Chen, Y., Basu, R., Gleghorn, M. L., Murakami, K. S., and Carey, P. R. (2011) Time-resolved events on the reaction pathway of transcript initiation by a single-subunit RNA polymerase: Raman crystallographic evidence. *J. Am. Chem. Soc.* **133,** 12544–12555

Structural Basis of Transcription Initiation by Bacterial RNA Polymerase Holoenzyme

Ritwika S. Basu, Brittany A. Warner, Vadim Molodtsov, Danil Pupov, Daria Esyunina, Carlos Fernández-Tornero, Andrey Kulbachinskiy and Katsuhiko S. Murakami

J. Biol. Chem. 2014, 289:24549-24559.
doi: 10.1074/jbc.M114.584037 originally published online June 27, 2014