THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 278, No. 14, Issue of April 4, pp. 12344–12355, 2003
Printed in U.S.A.

Co-overexpression of *Escherichia coli* RNA Polymerase Subunits Allows Isolation and Analysis of Mutant Enzymes Lacking Lineage-specific Sequence Insertions*

Received for publication, November 3, 2002, and in revised form, December 23, 2002
Published, JBC Papers in Press, January 2, 2003, DOI 10.1074/jbc.M211214200

Irina Artsimovitch‡, Vladimir Svetlov‡§, Katsuhiko S. Murakami¶, and Robert Landick‖**

From the ‡Department of Microbiology, Ohio State University, Columbus, Ohio 43210, the ¶Laboratory of Molecular Biophysics, Rockefeller University, New York, New York 10021, and the ‖Department of Bacteriology, University of Wisconsin, Madison, Wisconsin 53706

The study of mutant enzymes can reveal important details about the fundamental mechanism and regulation of RNA polymerase, the central enzyme of gene expression. However, such studies are complicated by the multisubunit structure of RNA polymerase and by its indispensability for cell growth. Previously, mutant RNA polymerases have been produced by *in vitro* assembly from isolated subunits or by *in vivo* assembly upon overexpression of a single mutant subunit. Both approaches can fail if the mutant subunit is toxic or incorrectly folded. Here we describe an alternative strategy, co-overexpression and *in vivo* assembly of RNA polymerase subunits, and apply this method to characterize the role of sequence insertions present in the *Escherichia coli* enzyme. We find that co-overexpression of its subunits allows assembly of an RNA polymerase lacking a 188-amino acid insertion in the β′ subunit. Based on experiments with this and other mutant *E. coli* enzymes with precisely excised sequence insertions, we report that the β′ sequence insertion and, to a lesser extent, an N-terminal β sequence insertion confer characteristic stability to the open initiation complex, frequency of abortive initiation, and pausing during transcript elongation relative to RNA polymerases, such as that from *Bacillus subtilis*, that lack the sequence insertions.

The amino acid sequences of the two largest subunits of cellular RNA polymerases (RNAPs),[1] called β′ and β in bacteria, are remarkably conserved among multisubunit RNAPs from eubacteria to the human homologues, RNAPI, RNAPII, and RNAPIII (1, 2). Recently determined crystal structures of yeast and bacterial RNAPs (3, 4) reveal the structural congruity responsible for this conservation, which extends far beyond the catalytically important residues (5–7). Representing the bulk of the enzyme, β′ and β (and their yeast homologues, Rpb1 and Rpb2) form two halves of a crab claw-shaped molecule, in which the secondary structures of the homologous subunits are nearly identical within 25 Å of the active site (located at the internal junction between the claws). The divergence among RNAP's large subunits generally increases toward the exposed surface of the molecule and frequently is manifest as insertions of up to several hundred amino acids that are characteristic of different evolutionary lineages (3). This pattern of surface sequence variation in RNAPs, which in eukaryotes includes additional subunits bound at the enzyme's periphery, led to the idea that the conserved catalytic core of RNAP is adapted to various environments and cellular milieus by the addition of surface modules that interact with different regulatory factors (4, 7, 8).

In bacteria, the most studied examples of lineage-specific sequence insertions (SIs)[2] occur in proteobacteria. In the proteobacteria, three easily recognizable insertions protrude from the surface of the enzyme (Figs. 1 and 6), two in β (SI1 between the conserved regions B and C and SI2 between regions G and H) and one in β′ (SI3 between regions G and H) (9, 10). These insertions are absent in Gram-positive bacteria and in *Deinococci*, from which the bacterial RNAP structures have been determined; the *Deinococci* contain a different SI between conserved regions A and B of β′ (3, 11). Both βSI1 and βSI2 tolerate insertions and partial deletions without affecting cell viability or *in vitro* RNAP activities (9, 12, 13). In contrast, partial deletions in β′SI3 are more deleterious; they reduce cell viability and confer defects in transcript cleavage and elongation, whereas the complete removal of β′SI3 inhibits assembly of the mutant β′ into core RNAP (14).

The observations that SIs can tolerate significant structural alterations without the loss of RNAP function led to them being described as dispensable (13–15). SIs in related bacteria exhibit somewhat greater variability than the sequences of full-length β and β′ (*e.g.* SIs are ~70% conserved between *Escherichia coli* and *Hemophilus influenzae* and ~60% conserved between *E. coli* and *Shewanella violacea*, whereas the remaining parts of β and β′ are ~85 and ~80% conserved between the two pairs, respectively). Nonetheless, their retention in distinct bacteria with significant sequence conservation suggests that they play some functional role. One role could be the recruit-

* This work was supported by National Institutes of Health Grant GM38660 and Department of Agriculture Grant WIS04022 (to R. L.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Supported by National Institutes of Health Grant GM 28575 (to Richard R. Burgess, Department of Oncology, University of Wisconsin, Madison, WI).

** To whom correspondence should be addressed. Fax: 608-262-9865; E-mail: landick@bact.wisc.edu.

[1] The abbreviations used are: RNAP, RNA polymerase; SI, lineage-specific sequence insertion; ORF, open reading frame; CBP, chitin-binding protein; nt, nucleotide(s); aa, amino acid(s); IPTG, isopropyl-1-thio-β-D-galactopyranoside; DTT, dithiothreitol.

[2] Sequence insertions in *E. coli* RNAP previously have been designated DR, to signify dispensable regions (see Ref. 7 and references therein). We prefer the less restrictive abbreviation SI to signify lineage-specific sequence insertions, because *E. coli* cells containing SI deletions are inviable in at least some conditions and exhibit altered biochemical properties, suggesting that these features of RNAPs play important functional roles. Further, the SI nomenclature is applicable to other bacterial RNAPs for which functional tests have not been performed.

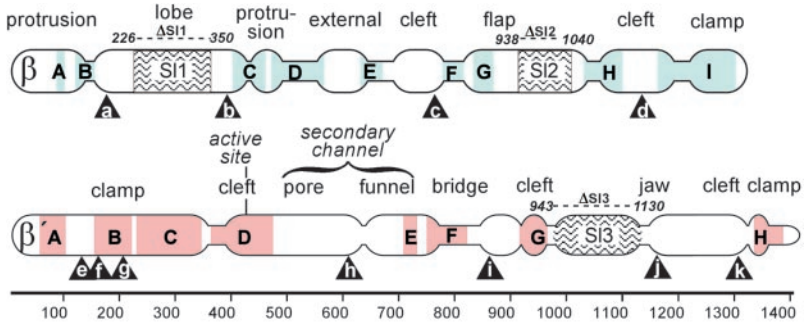

FIG. 1. **Linear maps of the β′ and β subunits of RNAP.** Conserved regions of β′ (*A–H*) and β (*A–I*) are shown in *pink* and *blue*, respectively, with the structural domains of RNAP indicated by the wider segments of the subunit maps and labeled using the nomenclature of Darst and Kornberg (3, 4). SI indicates the positions of amino acid sequence insertions in *E. coli* RNAP that are not present in most bacterial species; respective boundaries are shown *above* each SI. The endpoints of deletions studied here are given at the *ends* of *dotted lines* (*numbers* listed are the last deleted residues). *Black triangles* marked with *letters* (*a–k*) indicate positions of other sequence insertions in various organisms. *a* (at β184), 178-aa insertion in *Odontella sinensis* chloroplast; *b* (at β391), 25-aa insertion in *Xanthomonas campestris*; *c* (at β774), 70-aa insertion in *Leptospira biflexa*; *d* (at β1135), 69-aa insertion in *Wolbachia pipientis*; *e* (at β′144), 290-aa insertion in *Thermotoga maritima*; *f* (at β′164), 115-aa insertion in *Thermus aquaticus*; *g* (at β′204), 115-aa insertion in *Aquifex aeolicus*; *h* (at β′618), 76-aa insertion in *Mycoplasma pulmonis*; *i* (at β′869), 99-aa insertion in *T. maritima*; *j* (at β′1161), 53-aa insertion in *A. aeolicus*; *k* (at β′ 1302), 68-aa insertion in *T. maritima*. These sequence insertions represent only a subset of a growing number of apparent lineage-specific insertions intercalated between the conserved regions of β and β′ subunits in bacterial RNAPs.

ment of cellular regulatory proteins. For instance, βSI1 appears to be the target of the phage T4 Alc termination factor (15). SI modules could also affect biochemical properties of RNAP more directly; both β′SI3 and βSI1 are proposed to make downstream DNA contacts that explain the greater stability and longer downstream footprints of *E. coli* RNAP open complexes (relative to, for instance, *Bacillus subtilis* RNAP) (7, 16).

We became interested in the sequence insertions in proteobacterial RNAP as possible explanations for the inability of *B. subtilis* RNAP to recognize hairpin-dependent pause signals as well as the decreased open complex longevity and decreased abortive initiation of *B. subtilis* RNAP (17). To facilitate study of RNAPs with precise SI deletions, we developed a polycistronic co-overexpression system for *E. coli* RNAP that relies on T7 RNAP-dependent transcription of all three core RNAP subunit genes (*rpoA*, *rpoB*, and *rpoC*) on a single plasmid. To facilitate purification of the recombinant RNAPs, we appended an intein-chitin-binding protein (CBP) module to the β′ C terminus and a hexahistidine-hemagglutinin tag to the N terminus of β. This system allowed us to purify and test precise deletions of all three SIs in *E. coli* RNAP and provides a generally useful method for the study of mutant *E. coli* RNAP enzymes that avoids the loss of activity and assembly problems that sometimes arise with *in vitro* reconstitution methods (18, 19).

EXPERIMENTAL PROCEDURES

Reagents and Proteins—Oligonucleotides (listed in Table I) were obtained from Operon Technologies (Alameda, CA); dNTPs were from U.S. Biochemical Corp.; NTPs were from Amersham Biosciences; [α-^{32}P]CTP was from PerkinElmer Life Sciences; and other chemicals were from Sigma. Restriction and modification enzymes were obtained from New England Biolabs. Linear DNA templates for *in vitro* transcription were generated by PCR amplification and purified using reagents (cat. A7170) from Promega (Madison, WI).

Construction of Deletion Mutants—To delete βSI1, an *rpoB* fragment was PCR-amplified from pRL702 with oligonucleotides 343 and 3135. The PCR product was digested with *Nru*I and cloned between two *Nru*I sites of pRL702. To delete βSI2 and β′SI3, site-directed PCR mutagenesis with two fully complementary oligonucleotides flanking the deleted fragment in pIA160 (β) or pRL663 (β′) was performed. Each oligonucleotide (see Table I) annealed on both sides of the deletion, forcing the intervening fragment to loop out. The shortened *rpoB* fragment located between the unique *Nco*I and *Cla*I sites was sequenced, excised, and recloned into *Nco*I, *Cla*I-cut pIA160 (pIA319) or *Nco*I, and *Cla*I-cut pIA178 (pIA302). The shortened *rpoC* fragment located between the unique *Sal*I and *Bsp*EI sites was sequenced, excised, and recloned into *Sal*I-, *Bsp*EI-cut pRL663. To obtain overexpression plasmids encoding β ΔSI1 and SI2, the mutant *rpoB* fragments were transferred to the T7 RNAP-based expression plasmid pIA423 (Fig. 2) on an *Nco*I to *Sda*I fragment. To obtain overexpression plasmids encoding β′ Δ(943–1130), the mutant *rpoC* fragments were transferred to the same expression plasmid on a *Bsm*I to *Xho*I fragment.

Design of the Polycistronic Overexpression Plasmid—The original plasmid for co-overexpression of *E. coli* RNAP subunit genes, pIA423 (Fig. 2; GenBank™ accession number AF533984), contains a polycistronic operon *rpoA-rpoB-rpoC**, flanked by a single T7 promoter and terminator sequences derived from pET21 (Novagen, Madison, WI); each ORF is preceded by a separate ribosome-binding site (Fig. 2). Introduction of a *Xho*I site at the 3′-end of the *rpoC* ORF resulted in the addition of two amino acids (LE) to the C terminus of the β′ subunit, whereas a new *Eag*I site in *rpoB* ORF is silent. In the absence of induction, expression is repressed by the product of the *lacI*^{q} gene (carried on the same plasmid) through *lacO* positioned upstream of *rpoA*. To facilitate purification of the recombinant RNAP, the *rpoC** fusion contains a CBP-intein module from pTYB3 plasmid (New England Biolabs) fused in-frame to the 3′-end of the *rpoC* ORF. Derivatives of pIA423 containing hexahistidine and a hemagglutinin epitope tag at the N terminus of β were prepared by ligation of *Nco*I-*Sda*I fragments from pIA160 and pIA178 (*rpoB*^{SF531}) into the corresponding sites of pIA423.

The co-overexpression plasmid pIA423 was created in the following steps (see also Table I). pET21α was constructed to express *rpoA* under the control of an IPTG-inducible T7 RNAP promoter. The *Hin*dIII site in the pET21α *rpoA* gene was eliminated without altering the encoded amino acid sequence by site-directed PCR mutagenesis with primers 3683 and 3684 to yield pIA287. pIA287 was converted to pIA299 first by introducing a new sequence between the *Bam*HI and *Nco*I sites downstream of *rpoA*. The resulting plasmid was then modified by insertion between its *Nco*I and *Hin*dIII sites of an *Nco*I to *Sbf*I fragment from pRW408 carrying most of *rpoB*, followed by an *Sbf*I to *Bsm*I fragment from pNF1346, carrying the remainder of *rpoB*, the *rpoBC* intergenic region, and the first half of *rpoC*, followed by a *Bsm*I to *Hin*dIII fragment from pRL663, carrying the remainder of *rpoC*, a C-terminal *Xho*I site, and a His_6 tag. pIA423 was then created by insertion between *Xho*I sites of pIA299 of a *Xho*I-*Sal*I-digested PCR product from pTYB3 (amplified with primers 3741 and 3742; Table I).

RNAP Purification—Plasmids encoding variants of *E. coli* RNAP (wild type and mutants) were transformed into BL21 λDE3 (20). A single colony was inoculated into 500 ml to 2 liters of LB + 100 μg of ampicillin/ml at 37 °C and grown until apparent A_{600} reached 0.3–0.5, at which point protein production was induced by the addition of IPTG to 1 mM. Cells were grown for 3 h at 37 °C, rapidly chilled on ice, collected by centrifugation for 15 min at 5000 × *g* and 4 °C, and resuspended in 50 ml of column buffer (20 mM Tris, pH 7.9, 5% glycerol, 500 mM NaCl, 1 mM EDTA). Protease inhibitor mixture (Sigma catalog no. P8465) was then added as recommended by the manufacturer, and the cells were disrupted by sonication. The resulting lysate was cleared by centrifugation for 20 min at 27,000 × *g* and 4 °C and then filtered through the 0.4-μm syringe filter (Nalgene). Chitin beads (5 ml; New

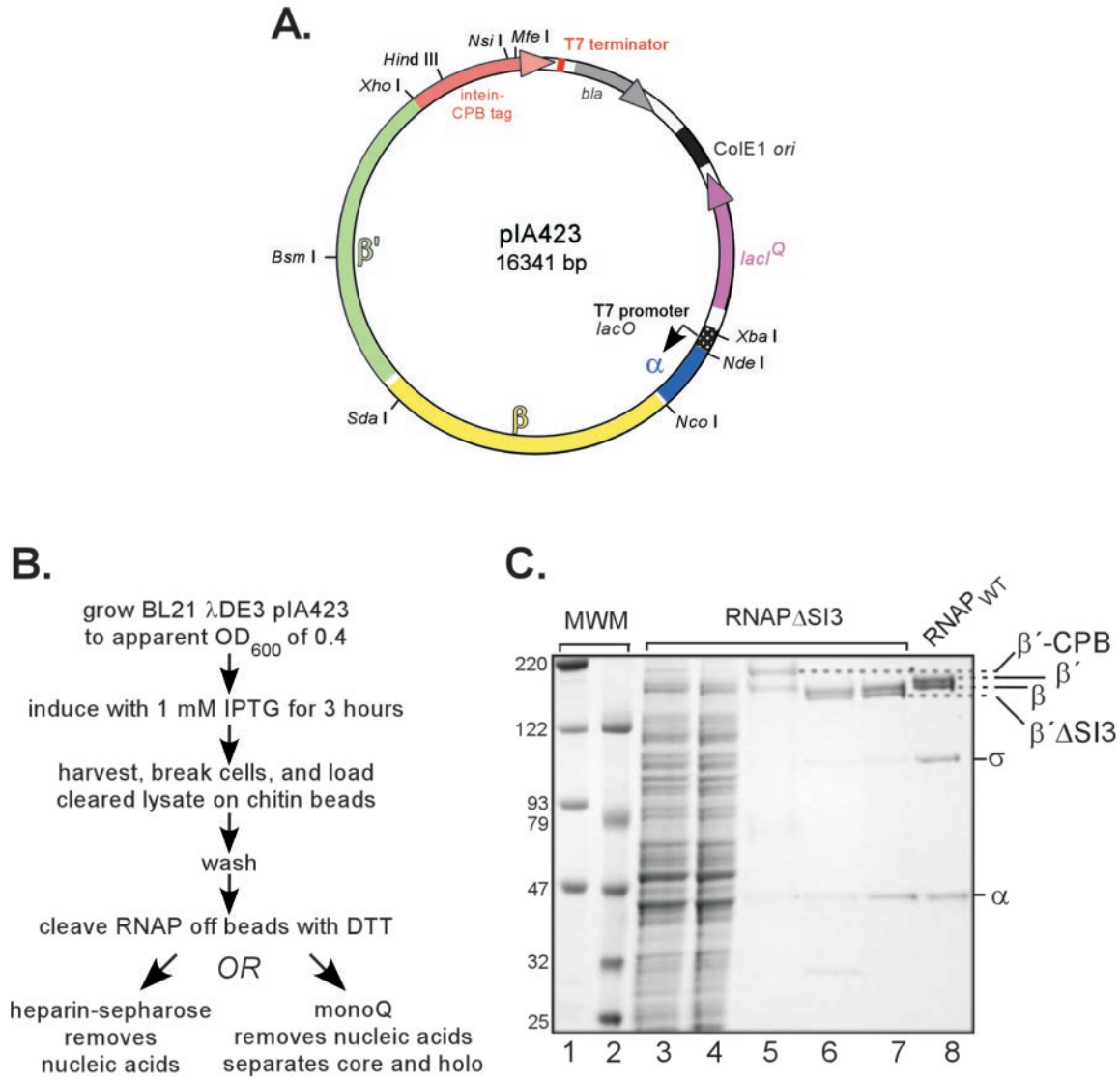

FIG. 2. **Purification of RNAP via the chitin protein binding tag.** *A*, overexpression plasmid pIA423 contains the *rpoA-rpoB-rpoC* gene cassette flanked by a single T7 gene 10 promoter and T7 terminator. The *rpoC* gene is fused in-frame to sequences encoding the CBP/intein module. *B*, purification scheme with times indicated. *C*, purification of $\beta'\Delta$SI3 mutant RNAP. 4–12% NuPAGE SDS-PAGE gel was loaded with (*lanes 1* and *2*) protein molecular weight standards (Bio-Rad) with sizes indicated on the *left*; *lane 3*, cleared lysate; *lane 4*, flow-through fraction; *lane 5*, bead fraction prepared by boiling the beads in SDS; *lane 6*, eluate from chitin column; *lane 7*, eluate from heparin column; *lane 8*, MRE600 holoenzyme prepared by the conventional method (27).

England Biolabs) were equilibrated with 10 volumes of column buffer in a 20-ml disposable column (Bio-Rad Econo-Pac), and cleared lysate was passed through the column by gravity flow, followed by 20 volumes of column buffer. To induce intein cleavage, the column was washed with 3 bed volumes of column buffer containing 50 mM DTT (to exchange buffer) and then incubated at 4–8 °C for 8–16 h (overnight). To elute the protein, column buffer (~4 ml) was added, and 0.2-ml fractions were then collected and tested for protein by Bradford assay and SDS-PAGE (4–12% NuPAGE gels; Invitrogen). Fractions containing RNAP were pooled, concentrated using Centriplus 100 or Ultrafree concentrators (Millipore Corp.) to 1–5 ml (depending on total RNAP recovered), and then dialyzed against loading buffer for heparin affinity or anion exchange chromatography. Chromatography was carried out using HiTrap columns and an Akta Prime low pressure chromatography system (Amersham Biosciences). For heparin affinity separation, samples were loaded onto the HiTrap Heparin HP column in 50 mM sodium phosphate (pH 6.9), 0.1 mM DTT buffer. For quaternary amine chromatography (Hi-Trap Q Sepharose Fast Flow), protein was loaded in 50 mM Tris-HCl (pH 8.0), 5% glycerol, 0.1 mM Na-EDTA, 0.1 mM DTT. Columns were washed with 10 column volumes of the loading buffer and eluted with a gradient (0–1.5 M) of NaCl in loading buffer over 20 column volumes. Elution peaks were identified by monitoring A_{280}, and their contents were further characterized using SDS-PAGE. Fractions containing RNAP were pooled and dialyzed against storage buffer (10 mM Tris, pH 7.9, 50% glycerol, 0.1 mM EDTA, 0.1 mM DTT, 0.1 M NaCl) for 12–14 h at 4 °C. Enzymes containing deletions in the β subunit were additionally purified by adsorption to Ni^{2+}-nitrilotriacetic acid beads (Qiagen), followed by washing and imidazole elution prior to dialysis against storage buffer. Typical yields from 2-liter cultures were 0.5–2 mg of purified RNAP, depending on properties of the particular enzyme. For the wild type RNAP, versions containing or lacking hexahistidine tags on β were purified and found to behave indistinguishably in the assays used here. *B. subtilis* core RNAP (21) and σ^{70} (22) were purified as described. Wild type and mutant RNAP holoenzyme (core $\beta'\beta\alpha_2$ plus σ^{70}) was prepared by incubating a 5-fold molar excess of σ^{70} with core enzyme for 30 min at 30 °C.

Open Complex Longevity Assays—Linear DNA template (40 nM) carrying the T7A1 promoter (pIA171) (23) was incubated with 50 nM RNAP holoenzyme (with σ^{70}) for 15 min at 37 °C in 50 μl of transcription buffer (20 mM Tris·HCl, 20 mM NaCl, 10 mM $MgCl_2$, 14 mM 2-mercaptoethanol, 0.1 mM EDTA, pH 7.9). Heparin was added to 14 μg/ml to sequester any free RNAP (at time 0); reaction aliquots were withdrawn at various times and combined with nucleotide substrates (ApU to 150 μM; ATP, GTP, and CTP to 2.5 μM; and 5 μCi of [α-^{32}P]CTP). Following a 10-min incubation at 37 °C, reactions were quenched by the addition of an equal volume of STOP buffer (10 M urea, 20 mM EDTA, 45 mM Tris borate, pH 8.3). For the zero time point, an aliquot was taken prior to heparin addition and incubated with substrates.

TABLE I
Strains, plasmids, DNA oligonucleotides, and transcription templates

Name	Description	Source or note
Strains		
BL21 λDE3	*E. coli* B F^- λ (T7 gene 1) *ompT gal[dcm]* [*lon*] *hsdS*$_B$(r_B^- m_B^-) λDE3	Refs. 20, 56
RL585	*leu*(Am) *trp*(Am) *lacZ* 2110(Am) *galK*(Am) *galE rsp*L *tsx sup*D43,74(Ts) *sue*A *rpoB*cI (Am) Δ(*rec*A-*srl*)306 *srl*-301::Tn10–84	Ref. 57
RL602	*leu*(Am) *trp*(Am) *lacZ* 2110(Am) *galK*(Am) *galE rpsL tsx sueB sueC supD43,74*(Ts) *rpoC325*(Am) Δ*(recA-srl)306 srl-301*::T*n*10–84	Ref. 40
DH5α	F′ *endA1 hsdR17* (r_k^- m_k^+) *glnV44 sup*E44 *thi-1 recA1 gyrA relA1* Δ(*lacIZYA-argF*)U169 *deoR* (ϕ80 *dlac*Δ(*lacZ*)M15)	
Plasmids		
pTYB3	Impact-CN™ system vector for the in-frame insertion into a polylinker upstream of an intein tag and CBP	New England Biolabs
pET21α	*Nde*I-*Bam*HI PCR fragment of *E. coli rpoA* ligated between *Nde*I and *Bam*HI of pET21c (Novagen) such that *rpoA* starts in the *Nde*I site and stops two nt (CC) before the *Bam*HI site.	This work
pRW408	*lacI*^{q}-P$_{trc}$ - *rpoB* - *bla* - ColEI *ori* plasmid	Ref. 58
pRL550	T7 A1 promoter-mutant initial transcribed sequence	Ref. 24
pRL663	*lacI*^{q}-P$_{trc}$ - *rpoC*His$_6$ - *bla* - ColEI *ori* plasmid	Ref. 59
pRL702	*lacI*^{q}-P$_{trc}$ - His$_6$-HA-*rpoB* - *bla* - ColEI *ori* plasmid. Insertion of an *Nco*I-, *Bsa*BI-digested 2195–106 PCR fragment from pRW408 into *Nco*I-, *Bsa*BI-digested pRW408. The N-terminal sequence of β from pRL702 is MAHHHHHHAYPYDVPDYAMVY (His$_6$ tag, hemagglutinin tag, and wild type β start are underlined).	This work
pIA150	Deletion of β aa 938–1040 (SI2) in pRL702	This work
pIA160	A silent *Xho*I site at β aa 998 in pRL702	This work
pIA171	T7 A1 promoter - A29 - *his* pause transcription template plasmid	Ref. 49
pIA178	β SF531 (*rif*^{d}18) in pIA160	This work
pIA287	*Hin*dIII site removed from pET21α with oligonucleotides 3653 and 3654	This work
pIA299	Expresses α + β + β′His$_6$ polypeptides from T7 promoter. Derived from pIA287 (see "Experimental Procedures").	This work
pIA301	Deletion of β′ aa 943–1130 (SI3) in pRL663	This work
pIA302	Deletion of β aa 226–350 (SI1) in pIA178	This work
pIA308	Deletions of β SI1 and SI2 in pIA178	This work
pIA311	Deletion of β SI2 in pIA178	This work
pIA319	Deletion β SI1 in pIA160	This work
pIA320	Deletions of β SI1 and SI2 in pIA160	This work
pIA329	Expresses α + β (His$_6$ *rif*^{d}18 ΔSI1) + β′-CBP/intein from T7 promoter	This work
pIA331	Expresses α + β + β′ΔSI3-CBP/intein from T7 promoter	This work
pIA332	Expresses α + β (His$_6$ *rif*^{d}18) + β′-CBP/intein from T7 promoter	This work
pIA333	Expresses α + β (His$_6$) + β′-CBP/intein from T7 promoter	This work
pIA334	Expresses α + β (His$_6$ *rif*^{d}18 ΔSI2) + β′-CBP/intein from T7 promoter	This work
pIA349	T7A1 promoter - *rfaQ ops* pause transcription template plasmid	Ref. 60
pIA423	Expresses α + β + β′-CBP/intein from T7 promoter	This work
pNF1346	*rplKAJL - rpoBC*(*rif*^{d}18) plasmid	Ref. 61
Oligonucleotides		
106	5′-CATCATGCGGTAGATTTCTA	To make pRL702
343	5′-GTACTCAGGTCAACTTTCTG	To create βΔSI2
645	5′-CAGTTCCCTACTCTCGCATG	T7A1 downstream
947	5′-GGAGAGACAACTTAAAGAG	T7A1 upstream
1454	5′-GGCTTTCTCATGCGTTCATG	T7A1 downstream
2195	5′-CACCCATGGCACACCATCACCACCATCACGCCTATCCATACGATGTGCCAG ATTATGCAATGGTTTACTCCTATACCGAGAAAAAACC	Used to insert His$_6$ and HA tags at *rpoB* N terminus
3135	5′-TAACTATCGCGACGATCTGGCACCGGGCGTG	To create βΔSI2
3177	5′-AGGATGATGGTGATGATGGTGG	Oligonucleotide competition experiments (Fig. 4*D*)
3653	5′-TGGCTGAACAACTGGAGGCTTTCGTTGACTTAC	To remove *Hin*dIII site in *rpoA*
3654	5′-GTAAGTCAACGAAGCCTCCAGTTGTTCAGCCA	To remove *Hin*dIII site in *rpoA*
3685	5′-GATCCCCGATCCGTCGACTTGTCAGCGAGCTGAGGAAACAGACCATGGA	*Nco*I-*Bam*HI linker for pIA299 construction
3686	5′-GATCTCCATGGTCTGTTTCCTCAGCTCGCTGACAAGTCGACGGATCGGG	*Nco*I-*Bam*HI linker for pIA299 construction
3691	5′-GTTCCACATCGGTGGTGCGGCATCTACCAAGGACATCACCGGTGGTCTGCC	To create β′SI3 deletion
3692	5′-GGCAGACCACCGGTGATGTCCTTGGTAGATGCCGCACCACCGATGTGGAAC	To create β′SI3 deletion
3693	5′-ACAGAGCAGATCCTCGACCTGTTCTTTTTACGTGTCGACCCAACTAACGACC	To create βSI1 deletion
3694	5′-GGTCGTTAGTTGGGTCGACACGTAAAAAGAACAGGTCGAGGATCTGCTCTGT	To create βSI1 deletion
3741	5′-GAATTCCTCGAGGGCTCTTCCT	To amplify intein-CBP tag from pTYB3
3742	5′-ATGCTGTCGACTCATTGAAGCTGCCACAAGG	To amplify intein-CBP tag from pTYB3
Transcription templates		
his pause	PCR of pIA171 with 947 and 645	
ops pause	PCR of pIA349 with 947 and 645	
Abortive initiation	PCR of pRL550 with 947 and 645	

Abortive Initiation Assays—Reactions were assembled on ice in 50 μl of transcription buffer with ApU at 150 μM, ATP and CTP at 20 μM, and 10 μCi of [α-^{32}P]CTP (3000 Ci/mmol). Linear pRL550 (24) DNA template was at 40 nM, and RNAP holoenzyme containing σ^{70} was at 50 nM. Transcription was initiated by shifting samples to 37 °C. Samples (5 μl) were removed at times indicated in the legend to Fig. 3*B* and after a final 5-min incubation with 500 μM each NTP (chase) and quenched by the addition of STOP buffer.

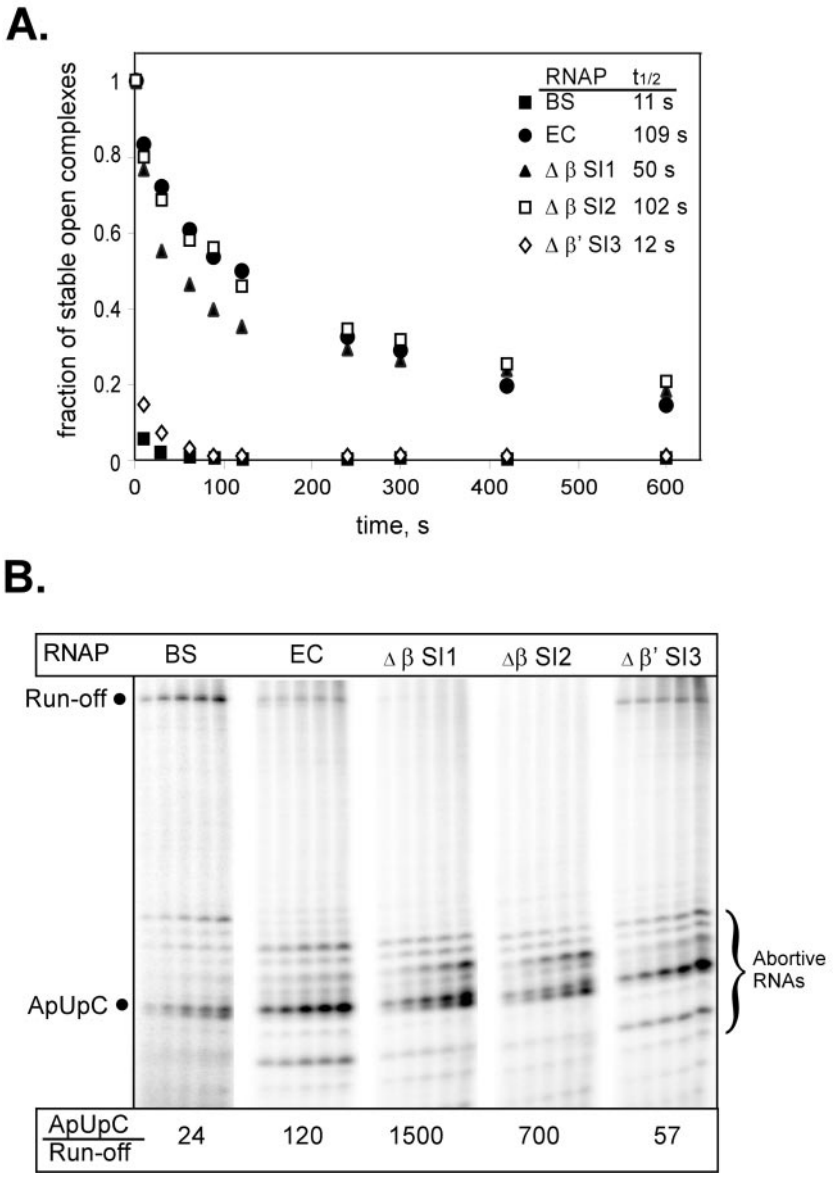

FIG. 3. **Transcription initiation of *E. coli*, *B. subtilis*, and ΔSI mutant RNAPs.** *A*, open complex stability *in vitro*. Plot of open complexes formed at T7 A1 promoter remaining as a function of time after the addition of heparin (see "Experimental Procedures"). *B. subtilis* (*BS*), *E. coli* (*EC*), and mutant holoenzymes lacking SIs were prepared with *E. coli* σ^{70} initiation factor. The half-lives of open complexes were estimated by nonlinear regression of the amount of transcript produced as a function of the time elapsed between the heparin addition and NTP addition to preformed open complexes. Relative to the amount of transcript produced by wild type RNAP with no heparin added, the other RNAPs yielded 0.2, 0.6, 0.5, and 1.1 equivalents of transcript in the absence of heparin (*B. subtilis*, Δβ SI1, Δβ SI2, and Δβ′SI3, respectively; adjusted to 1.0 for the zero time point). *B*, electrophoretic separation of [α-^{32}P]CTP-labeled RNAs formed on the pRL550 template as a function of time (during 5-, 10-, 20-, 30-, and 60-min incubation with substrates; see "Experimental Procedures"). Positions of the run-off RNA and the predominant abortive product, ApUpC, are indicated. The ratio of ApUpC trimer to the full-length RNA (corrected for the number of C residues) is shown *below* for each enzyme.

Single Round Pause Assays—Linear DNA template encoding the *his* pause signal was generated by PCR amplification of pIA171 (23). Halted A29 elongation complexes were formed during a 15-min incubation of 40 nM DNA template and 50 nM RNAP holoenzyme at 37 °C in 50 μl of transcription buffer in the absence of UTP, with ApU at 150 μM, ATP and GTP at 2.5 μM, and CTP at 1 μM, with ^{32}P derived from [α-^{32}P]CTP (3000 Ci/mmol). Transcription was restarted by the addition of 20 μM GTP; 150 μM ATP, UTP, and CTP; and heparin to 100 μg/ml. Samples were removed at the times listed in the figure legends and after a final 5-min incubation with 250 μM each NTP (chase) and were quenched as above.

Sample Analysis—Samples were heated for 2 min at 90 °C and separated by electrophoresis in denaturing acrylamide (19:1) gels (7 M urea, 0.5× TBE; 8% for pause assays, 15% for abortive initiation and open complex stability assays). RNA products were visualized and quantified using a PhosphorImager and ImageQuant Software (Amersham Biosciences). Pause half-life (the time during which half of the complexes reenter the elongation pathway) was determined by nonlinear regression analysis as described previously (25).

RESULTS

Co-overexpression of RNAP Subunits Allows Study of the Role of Proteobacterial Sequence Insertions—To allow facile expression, assembly, and purification of poorly assembled or toxic mutant RNAPs, such as β′ΔSI3, we designed a vector that expresses α, β, and β′ polypeptides from a single T7 promoter; the β′ subunit is fused to the CBP tag (see "Experimental Procedures"). We predicted that expressing the mutant core RNAP from this plasmid would facilitate its assembly *in vivo* and would allow purification via the CBP tag (Fig. 2). This approach has several advantages. First, only the fully assembled core ($\alpha_2\beta\beta'$) RNAP is purified because β′ recruitment is the last step in the assembly pathway (26) and because the expression levels from this vector follow the assembly pathway of RNAP (α>β>β′; data not shown). Second, intein-mediated cleavage removes the CBP tag to release the assembled RNAP from the matrix; therefore, the purified enzyme does not carry additional protein segments. Third, the entire purification can be completed quickly, is relatively inexpensive, and yields RNAP that is sufficiently pure for *in vitro* transcription (Fig. 2). Fourth, the plasmid-encoded β and β′ subunits assemble preferentially with each other (and not with the chromosomal subunits; data not shown), allowing combination of substitutions in different subunits and more homogenous population of purified RNAPs. To purify RNAP with altered β subunit, we also constructed a version of the co-overexpression plasmid with a hexahistidine tag at the N terminus of β (see Table I).

Using the co-overexpression plasmid, we were able to obtain

active Δβ′SI3 RNAP (Fig. 2) as well as RNAPs with precise excisions of βSI1 or βSI2. We found that RNAP eluted directly from chitin beads was suitable for *in vitro* transcription but sometimes exhibited reduced activity relative to RNAP purified by conventional methods (27). We hypothesized that this low activity arose from residual binding of nucleic acids to the RNAPs; the addition of one chromatography step on heparin- or quarternary amine-Sepharose yielded overexpressed RNAP of an activity and purity comparable with that obtained by conventional purification (Fig. 2 and data not shown).

When the β′ΔSI3-intein-CBP fusion was expressed from a T7 promoter plasmid that did not encode the other RNAP subunits, essentially all tagged β′ was found in inclusion bodies (data not shown). Therefore, co-overexpression with α, β, and β′ subunits facilitates RNAP assembly. Perhaps translation of β and β′ from the same mRNA facilitates proper interaction, or the elevated level of the initially formed $\alpha_2\beta$ complex simply allows plasmid-encoded β′ to compete effectively with chromosomally encoded β′ for assembly.

β′SI3 Stabilizes Open Initiation Complexes—Using RNAPs with precise SI deletions obtained by co-overexpression, we first tested the effects of the *E. coli* SIs on open initiation complexes. Fully mature *E. coli* open complexes, which form on many but not all cellular promoters, exhibit extended contacts of RNAP with DNA from ~−55 to ~+20 relative to the transcription start site and melting of the DNA duplex between ~−12 and ~+2. These *E. coli* open complexes are long lived and collapse back to closed complexes at rates of $<0.01\ s^{-1}$ (28–31). In contrast, *B. subtilis* RNAP, which lacks insertion sequences, forms open complexes that are in rapid equilibrium with closed promoter complexes (32, 33). In *B. subtilis* RNAP open complexes, the DNA downstream of the start site is not efficiently protected against DNase I digestion (32, 34), and the melted region is shortened in the downstream direction (35). Interestingly, a deletion variant of the *E. coli* β subunit that includes part of βSI1 but extends beyond its boundaries (Δ186–433) forms open complexes that (at 37 °C) exhibit a melting pattern similar to those formed by the *B. subtilis* enzyme (35).

To measure the relative stability of open complexes formed by wild type or ΔSI RNAPs, we challenged open complexes formed on the T7 A1 promoter with the polyanion heparin. Although for many promoters heparin binds only free RNAP that is in equilibrium with closed initiation complexes and has no effect on open complexes, heparin directly attacks open complexes formed at the T7 A1 promoter and displaces RNAP from the promoter DNA in addition to binding free RNAP (36). Upon heparin addition to 14 μg/ml, transcriptionally competent open complexes (as assayed by the addition of NTPs and formation of an RNA transcript at different time intervals following heparin challenge) disappeared at a pseudo-first order rate of $0.006\ s^{-1}$ for wild type and ΔβSI2 RNAP, about twice as fast for ΔβSI1 RNAP, and about 10 times faster for Δβ′SI3 RNAP and *B. subtilis* RNAP (Fig. 3*A*). We conclude that the characteristic instability of *B. subtilis* open complexes is mimicked by Δβ′SI3 *E. coli* RNAP and that βSI1, but not βSI2, also contributes to the stability of *E. coli* open complexes.

β′SI3 Promotes Abortive Initiation—A second property that distinguishes *B. subtilis* and *E. coli* RNAPs is an ability of *B. subtilis* RNAP to escape promoters at which *E. coli* RNAP is trapped in abortive synthesis. We found this previously using a variant of the T7 A1 promoter whose initially transcribed sequence (5′-AUCCCACACC . . . *versus* wild type 5′-AUCGAGAGGG . . .) appears to act cooperatively with strong contacts of *E. coli* RNAP to promoter DNA to trap the enzyme in abortive synthesis (24); *B. subtilis* RNAP is able to escape from the mutant T7 A1 promoter with relatively few abortive products made, and the patterns of abortive products also differ between the *B. subtilis* and *E. coli* enzymes (17).

To investigate whether any of the sequence insertions in *E. coli* RNAP might explain its characteristically poor escape from the mutant T7 A1 promoter, we measured the abortive to productive RNA product ratios for the various RNAPs on this promoter (Fig. 3*B*). Interestingly, deletion of β′SI3 reduced the abortive to productive product ratio, although only to a level still ~2-fold greater than found for *B. subtilis* RNAP. However, deletion of either SI in β dramatically increased the ratio. Thus, the *E. coli* sequence insertions profoundly affect promoter escape, although in different directions; βSI1 and βSI2 promote escape, whereas β′SI3 inhibits escape.

The patterns of abortive products produced also are revealing (Fig. 3*B*). Both ΔβSI1 and ΔβSI2 RNAPs produce an altered pattern of abortive products relative to wild type or β′ΔSI3 RNAPs (compare the 60-min *lanes* for each RNAP in Fig. 3*B*). However, although β′ΔSI3 RNAP exhibited increased promoter escape, like *B. subtilis* RNAP, it does not recapitulate the distinctive pattern of *B. subtilis* abortive products. We conclude that the absence of the Proteobacterial sequence insertions substantially, although not completely, accounts for the distinctive initiation properties of *B. subtilis* RNAP.

β′SI3 Slows Escape from Pause Sites—Another key difference between the *E. coli* and *B. subtilis* RNAPs is their responses to certain hairpin-dependent pause sites. Both enzymes can recognize hairpin-independent pause signals as well as a *B. subtilis* P RNA pause site at which a nascent RNA hairpin 12 nt upstream from the pause is important in the presence, but not in the absence, of NusA (17). However, *B. subtilis* RNAP fails to recognize the *his* pause site, at which a nascent RNA hairpin 11 nt upstream from the pause site contributes a factor of 5–10 to the delay in pause escape through an interaction with the β flap domain of the *E. coli* RNAP (17, 37).

The structural basis for this difference is not known; *B. subtilis* RNAP might be unable to respond to the RNA hairpin formation or to other components of this signal (the downstream DNA, the 3′-proximal RNA, and the nucleotides in the active site) that also slow pause escape independently of the RNA hairpin. We previously found that replacement of the *E. coli* β flap with the corresponding fragment of the *B. subtilis* β did not alter pausing (17), suggesting that the difference in their response to the *his* pause signal lay elsewhere. Thus, the sequence insertions in *E. coli* RNAP were the next logical possibility.

To ask if the *E. coli* sequence insertions play a role in pause site recognition, we first tested pausing on the *his* pause template. We found that, unlike *B. subtilis* RNAP, ΔβSI1, ΔβSI2, and Δβ′SI3 RNAPs all recognized the *his* pause site (Fig. 4). However, Δβ′SI3 RNAPs escaped from the pause site ~5 times faster than wild type and exhibited an ~2-fold decrease in the efficiency of pause site recognition. To ask if the residual pausing by Δβ′SI3 RNAPs was still hairpin-dependent, we tested the effect of an antisense oligonucleotide that pairs to nascent pause RNA including the two 5′-most nt of the pause hairpin stem (Fig. 4*A*) and that eliminates the 5–10-fold contribution of the hairpin to pausing by disrupting its structure (23). If β′SI3 somehow mediated the effect of the pause hairpin, despite their locations on opposite sides of the paused TEC, we would expect the antisense oligonucleotide to have little effect on pausing by Δβ′SI3 RNAP. However, the antisense oligonucleotide reduced the pause half-life of Δβ′SI3 RNAP by a factor of ~3 (Fig. 4*D*), suggesting that β′SI3 contributes to pausing through interactions that are largely independent of the hairpin interaction. We verified this conclusion by testing the effect of Δβ′SI3 on

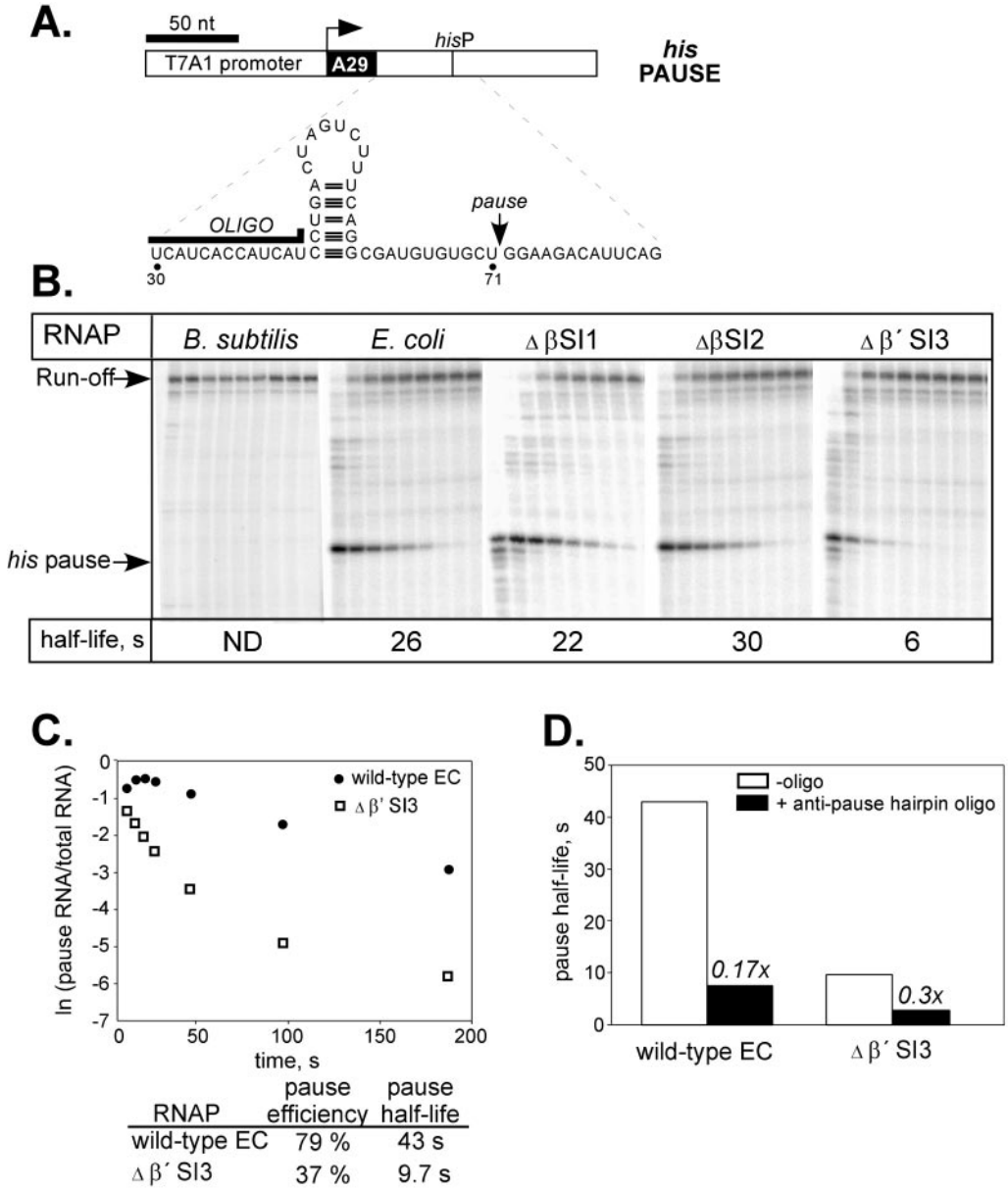

FIG. 4. **Pausing at the *his* pause site.** *A*, schematic of *his* pause template showing sequence in the region of the *his* pause. *B*, electrophoretic separation of RNAs formed on the *his* pause template as a function of time. For each *panel*, samples were taken at 5, 10, 20, 30, 45, 60, 120, and 240 s; the *last lane* in each *panel* (chase) contains sample that was incubated for an additional 5 min with 250 μM each NTP after completion of the time course. Preformed [α-^{32}P]CMP-labeled A29 complexes were incubated with 20 μM GTP, 150 μM ATP, CTP, and UTP on the *his* pause template. Prominent transcripts are pause RNA transcript (*his*P; 71 nt) and run-off RNA transcript. *C*, a plot of pause RNA fraction *versus* time for the wild type *E. coli* RNAP and β'ΔSI3 enzyme. The reactions were performed as in *A*, except that the chase was at 10 μM GTP to allow for more accurate measurements of the pause efficiency and half-life. *D*, effect of the antisense oligonucleotide (shown in *A*) competition on the half-life of *his* paused transcription complexes.

pausing at the hairpin-independent *ops* pause site, where the deletion had effects on half-life and efficiency nearly identical to the effects observed at the *his* pause site (Fig. 5). However, Δβ'SI3 did not fully recapitulate the pausing behavior of *B. subtilis* RNAP, which bypasses the *his* pause site without apparent delay (Fig. 4*B*) and transcribes through the *ops* pause site even more rapidly than does Δβ'SI3 RNAP (17).

We conclude that β'SI3 plays a central role in the strong pausing behavior of *E. coli* RNAP but that differences between *E. coli* and *B. subtilis* RNAP in addition to β'SI3 must contribute to reduced pausing by the latter enzyme. Interestingly, a partial deletion in β'SI3 (Δ1091–1130) displays the opposite effect on pausing as the precise β'SI3 deletion that we studied; *i.e.* it increases rather than decreases pausing (14). We return to the implications of this discrepancy under "Discussion."

Multiple Deletions of E. coli RNAP Sequence Insertions—Given the distinct effects of βSI1 and β'SI3 on open complex longevity, abortive initiation, and pausing, we wondered how these effects would combine in an enzyme lacking both SIs. All attempts to produce such an enzyme failed, however (data not shown). Derivatives of pIA423 that expressed βΔSI1 and β'ΔSI3 never yielded assembled RNAPs bearing the deletions, even when we included on the co-expression plasmid *rpoZ*, the gene encoding the dispensable RNAP subunit ω that is reported to promote RNAP assembly (38).

Deletions of the Sequence Insertions Impair Growth of Bacteria—Finally, we wished to test the dispensability of sequence insertions in RNAP for bacterial growth. Studies conducted prior to the availability of high resolution RNAP structures on RNAPs lacking portions of βSI1, βSI2, or β'SI3 led to the idea

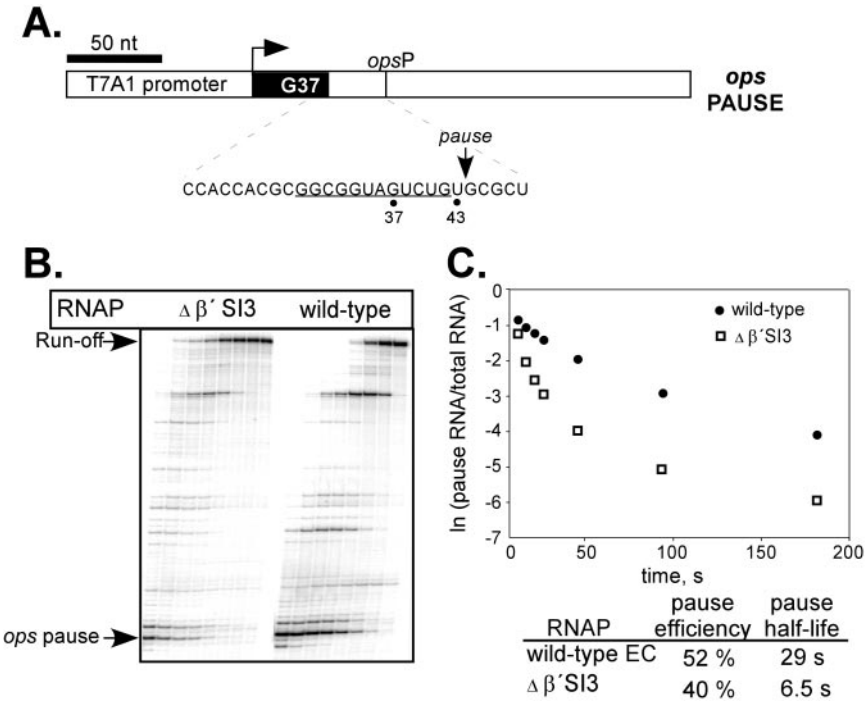

FIG. 5. **Pausing at the *ops* pause site.** *A*, schematic of *ops* pause (underlined) template showing sequence in region of the *ops* pause and electrophoretic separation of RNAs formed on the *ops* pause template as a function of time. Preformed [α-^{32}P]CMP-labeled G37 complexes were allowed to elongate at 10 μM GTP, 150 μM ATP, CTP, and UTP, and the samples were taken at 5, 10, 15, 20, 40, 90, and 180 s. The *last lane* in each *panel* contains chase sample that was treated as described in the legend to Fig. 4. Prominent transcripts are pause RNA transcript (*ops*P; 43 nt) and run-off RNA transcript (*RO*). *B*, the fractions of paused RNA were plotted against time; pause half-lives and pause efficiencies were determined as described under "Experimental Procedures" and in Ref. 25. *C*, a plot of pause RNA fraction *versus* time for the wild-type *E. coli* RNAP and $\beta'\Delta$SI3 enzyme determined from the samples shown in *panel A*.

that these were dispensable regions, at least for core RNAP functions (9, 12–14, 39). Now that we could define precise deletions of the sequence insertions and express functional enzymes containing these deletions *in vivo*, we wished to revisit this idea and ask to what extent the sequence insertions are required for bacterial growth.

To test the requirement for sequence insertions for bacterial growth, we expressed the wild type or ΔSI subunits from plasmids on which their expression either singly or in combination with the other RNAP subunits could be regulated by IPTG and *lac* repressor encoded on the same plasmids (Table I). We transformed the plasmids into strains that express wild type RNAP from the chromosome (rich medium, wild type subunit from chromosome expressed) (Table II) or in which expression of the corresponding subunit from the chromosome could be inactivated at 39 or 42 °C (for *rpoC* and *rpoB*, respectively; wild type subunit from chromosome not expressed, Table II).

We found that plasmid-encoded expression of mutant β subunit lacking SI1 or SI2 in the presence of wild type β encoded on the chromosome had little effect on cell growth However, $\beta\Delta$SI1 was unable to substitute for wild type β at 42 °C, and it compromised growth at 30 °C when it contained a rifampicin-resistant amino acid substitution (SF531) not present in the chromosomal copy of *rpoB* and when rifampicin was added to the medium (Table II). Growth on rifampicin at 30 °C was not restored when $\beta^{SF531}\Delta$SI1 was co-expressed with β' and α (conditions that yielded assembled $\beta\Delta$SI1 for purification). Thus, $\beta\Delta$SI1 appears to be incorporated into RNAP but unable to support cell growth. $\beta\Delta$SI2 was able to support growth on rich medium even at 42 °C or in the presence of rifampicin at 30 °C when combined with SF531. However, $\beta^{SF531}\Delta$SI2 was unable to support growth on minimal medium containing rifampicin, when expressed either singly or together with α and β'. Thus, both sequence insertions in *E. coli* β appear necessary for growth on minimal medium, suggesting that they play some regulatory role in RNAP function that becomes essential in stringent growth conditions.

We found that β'SI3 also was required for growth in some conditions; however, the requirements for this sequence insertion are complex. A $\beta'\Delta$SI3 plasmid could not be transformed into RL602 (Table I) in which expression of chromosomally encoded β' is temperature-sensitive, even in the absence of IPTG (probably because low, uninduced expression of $\beta'\Delta$SI3 produced too much defective enzyme when combined with the lowered expression of β' in RL602). The $\beta'\Delta$SI3 co-expression plasmid could be transformed into the conditional *rpoC* strain but blocked growth when induced at the permissive temperature and failed to support growth at the nonpermissive temperature. Although induction of $\beta'\Delta$SI3 in a strain producing wild type β' lowered plating efficiency by 10^2, it actually raised plating efficiency by $>10^3$ relative to wild type β', when plated in the presence of an antibiotic (microcin J25) that inhibits RNAP upon binding at a site near the β'SI3 in the RNAP secondary channel (62). Co-expression $\beta'\Delta$SI3 with β and α was not inhibitory in a strain expressing wild type RNAP (Table I) and conferred nearly 100% plating efficiency in the presence of microcin J25 (62). However, the growth rate of these colonies in the presence of microcin J25 is drastically slower than strains carrying the prototypical microcin J25 resistance mutation (*rpoC*^{T931I}). On balance, it appears that β'SI3 is essential for normal growth of *E. coli* but that, at least in the presence of microcin J25, RNAP lacking β'SI3 can still support slow growth on rich medium.

We conclude that sequence insertions, although dispensable for basic RNAP function at all stages of the transcription cycle, nonetheless are important to growth of the bacterium. This may reflect effects of the sequence insertions on the basic steps in transcription by RNAP that become crucial for expression of

TABLE II
Effect of expressing plasmid-encoded RNAP subunits lacking sequence insertions on the growth of E. coli

Plasmid-encoded subunits	Efficiency of Plating[a]			
	Rich medium, WT subunit from chromosome expressed[b]	Rich medium, WT subunit from chromosome not expressed[c]	Rich medium, WT subunit expressed; inhibitor present[d]	Minimal medium, WT subunit expressed; inhibitor present[e]
β subunit analysis				
WT β	0.9	1	$<10^{-6}$	$\leq 10^{-6}$
$\beta\Delta$SI1	1	$\leq 10^{-6}$	ND[f]	ND
$\beta\Delta$SI2	1	0.5	ND	ND
β^{SF531}	1	0.1	1	1
$\beta^{SF531}\Delta$SI1	0.1	ND	0.01	$\leq 10^{-6}$
$\beta^{SF531}\Delta$SI2	0.5	ND	0.5[f]	$\leq 10^{-6}$
$\beta'\beta\alpha_2$	1	1	$\leq 10^{-6}$	$\leq 10^{-6}$
$\beta'\beta^{SF531}\alpha_2$	1	0.1	1	1
$\beta'\beta^{SF531}\Delta$SI1$\alpha_2$	1	ND	0.006	$\leq 10^{-6}$
$\beta'\beta^{SF531}\Delta$SI2$\alpha_2$	0.9	ND	0.5[g]	$\leq 10^{-6}$
β' subunit analysis				
WT β'	1[h]	1	$\leq 10^{-5}$[h]	ND
$\beta'\Delta$SI3	0.02	ND	0.01[i]	ND
β'^{T9311}	1[h]	ND	0.8[h]	ND
$\beta'\beta\alpha_2$	1[h]	1	$\leq 10^{-5}$[h]	ND
$\beta'\Delta$SI3$\beta\alpha_2$	0.6[h,i]	$\leq 10^{-5}$	0.5[h,i]	ND

[a] RNAP subunits were expressed singly from pRL702 (β) or pRL663 (β'), or co-expressed from pIA423 ($\beta'\beta\alpha_2$) derivative plasmids (see Table I). To measure plating efficiencies, plasmids were transformed into the indicated strains, grown to A_{600} of ~0.3, an incubated with 1 mM IPTG for 1 h to induce expression of WT or mutant subunit. Serial dilutions of these cultures were then plated on agar medium containing 100 μg of ampicillin/ml, with or without 1 mM IPTG for calculation of plating efficiencies. All plating efficiencies are the average of quadruplicate measurements among which the S.E. was <30%. Although the coexpression plasmid ordinarily depends on T7 RNAP for overexpression of RNAP, the amount of RNAP generated from transcripts that initiate at other *E. coli* RNAP promoters on the plasmid is sufficient to support growth of an *E. coli* strain not containing T7 RNAP, provided that IPTG is present to remove *lac* repressor from the operator site at the beginning of *rpoA* on the plasmid.

[b] Plating efficiency = (colonies formed with IPTG/colonies formed without ITPG) in strain RL585 at 30 °C for the β subunit and strain DH5α at 37 °C for the β' subunit (see Table 1 and "Experimental Procedures"). Under these conditions, the WT β and β' subunits are present.

[c] Plating efficiency = (colonies formed at 42 °C with IPTG/colonies formed at 30 °C without IPTG) in strain RL585. At 42 °C, the RL585 chromosomal copy of β is inactivated (57). Plating efficiency when WT β' was absent could not be determined, because the strain that conditionally expresses β' (RL602) (40) could not be transformed with plasmids encoding $\beta'\Delta$SI3 even under noninducing conditions. Plating efficiencies for the co-overexpression plasmids were determined using derivatives of strains RL585 and RL602 in which λDE3 was integrated in their chromosomes.

[d,e] Plating efficiency on antibiotic (inhibitor) plates = (colonies formed with IPTG and antibiotic colonies formed with IPTG but without antibiotic). β mutants (top part) were grown in RL585 (*d*) or DH5α (*e*) at 30 °C with 25 μg of rifampicin/ml. β' mutants (bottom part) were grown in DH5α at 37 °C with 40 μg microcin J25/ml.

[f] ND, not determined.

[g] RL585 transformed with $\beta\Delta$SI2 plasmid formed small colonies (~⅕ of WT in diameter) in the absence of the WT β (at 42 °C or in the presence of rifampicin).

[h] From Ref. 62.

[i] DH5α transformed with $\beta'\Delta$SI3 or $\beta'\Delta$SI3$\beta\alpha_2$ plasmid formed tiny colonies (~1⁄10 of β'^{T9311} in diameter) visible only after prolonged incubation (>40 h) at 37 °C in the presence of IPTG.

certain genes or in some growth conditions but could also reflect roles in yet undescribed but essential interactions of transcription factors similar to the previously described interaction of the T4 Alc protein with βSI1 (15).

DISCUSSION

We have described a co-overexpression system for mutant *E. coli* RNAPs and its use to investigate whether sequence insertions in the β and β' subunits could explain the different initiation and elongation properties of *E. coli* RNAP compared with *B. subtilis* RNAP, which lacks these insertions. We found that co-overexpression facilitates assembly of otherwise difficult to assemble or toxic mutant RNAPs and that the sequence insertions, most notably β'SI3, confer some of its distinctive biochemical properties on *E. coli* RNAP and partially account for its differences from *B. subtilis* RNAP. We will discuss the implications of these findings for the study of mutant RNAPs, the function of the *E. coli* sequence insertions with focus on β'SI3, and the evolution of sequence insertions in RNAP.

An RNAP Overexpression System That Promotes Efficient Assembly—Previous studies of toxic RNAP mutants have relied either on conditional expression of a tagged mutant subunit followed by tag affinity separation from wild type RNAP (or tagging and removal of the wild type enzyme from the RNAP preparation) (40) or *in vitro* reconstitution of RNAP by gradual renaturation of mixtures of the denatured subunits (19, 41–43). $\beta'\Delta$SI3 illustrates the limitation of these approaches. When expressed as an individual subunit *in vivo*, $\beta'\Delta$SI3 fails to out-compete wild type β' for assembly into useful amounts of mutant enzyme; when used in *in vitro* reconstitution protocols, it fails to assemble into RNAP (see "Results" and Ref. 14).

Co-overexpression of wild type and mutant tagged subunits solved these problems for $\beta'\Delta$SI3 and has allowed us to isolate substantial quantities of many mutant RNAPs, including notably toxic mutants such as the $\beta\Delta$(900–909) flap mutant RNAP (37). Thus, the co-overexpression plasmid provides a powerful tool for genetic and biochemical analysis of RNAP. The resulting enzymes are substantially pure after two steps, primarily because chitin affinity chromatography is much more selective than the hexahistidine-Ni^{2+}-nitrilotriacetic acid-agarose method that has been widely used to date (44), and has the advantage of being assembled by the normal pathway *in vivo*. We anticipate that the co-overexpression plasmid will also prove useful for selection of RNAP suppressor mutants of inviable amino acid substitutions in the enzyme. Finally, because RNAP is pure enough for *in vitro* transcription assays after just the single chitin column step, the method should facilitate rapid screening of sets of closely related alterations to RNAP.

In using the co-overexpression method for mutant RNAP assembly and purification, we encountered a few problems that require care to avoid. First, the chitin matrix apparently can break down when batch binding and elution are used, which prevents retention of the tagged enzyme on the solid chitin

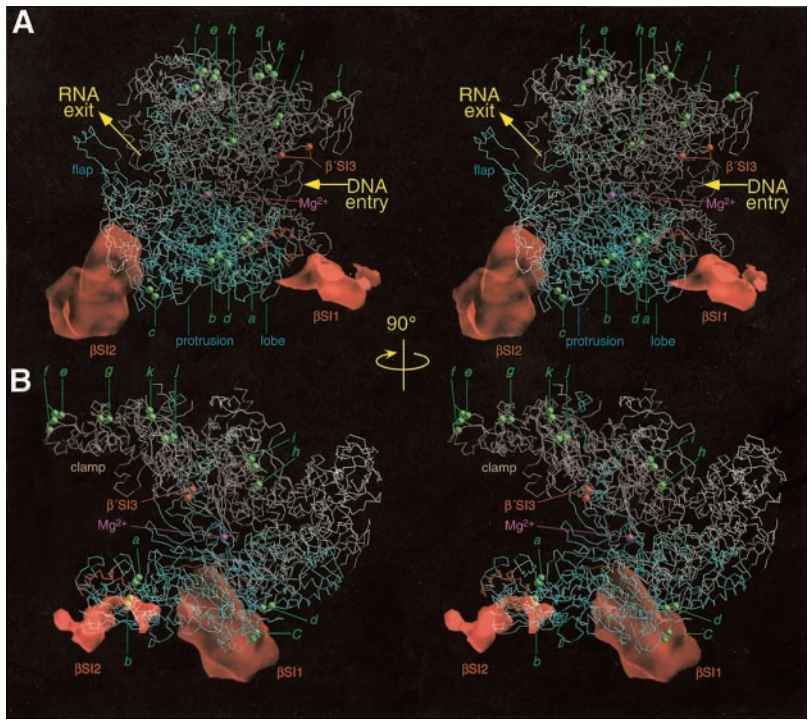

FIG. 6. **Locations of sequence insertions in *E. coli* RNAP and in other bacterial RNAPs.** *E. coli* RNAP is depicted in paired stereo images as observed by electron crystallography by Darst *et al.* (10) using the atomic coordinates of *T. aquaticus* RNAP adjusted to match the electron density of *E. coli* RNAP (the main channel of RNAP is more open in the *E. coli* structure than in the *T. aquaticus* structure). *Pink*, β′; *cyan*, β; *white*, α and ω; *magenta*, Mg^{2+} at catalytic center. (Note that the ω subunit is included based on the *T. aquaticus* RNAP structure; it was not seen in the *E. coli* RNAP electron microscopy structure.) The locations of βSI1 and βSI2 are depicted in red space fill (the density for βSI1 represents only a portion of the space expected to be occupied by this insertion; the remainder is disordered in the electron crystal structure (10)). The location of β′SI3, which is completely disordered and not visible in the electron crystal structure, is depicted based on its junction to the β′G loop as *red balls*. The locations of sequence insertions shown in Fig. 1 are marked by *green letters* and lines. *A*, view into the main channel of RNAP with the enzyme oriented for transcription from *left* to *right*. *B*, view into the NTP entry channel of RNAP.

matrix. We used slow passage of lysates through a column of chitin matrix to avoid this problem. Second, the efficiency of DTT-mediated cleavage of the *Sce* VMA intein that connects β′ to the CBP was usually less than 100%, and both the binding capacity and the cleavage efficiency appeared to vary among lots of chitin matrix. It is advisable to test new batches of matrix before committing valuable samples; some loss of material due to inefficient cleavage appears unavoidable. Third, we sometimes observed loss of expression of one or more of the RNAP subunits when the co-expression plasmid was maintained in *recA*$^+$, T7 RNAP-expressing strains, apparently due to plasmid rearrangements, mutations, or recombination with the chromosomal copies of *rpoA*, *rpoB*, or *rpoC*. The last problem was substantially eliminated by maintaining the plasmids in *recA* strains such as DH5α except during expression. Finally, we emphasize that not all RNAP mutants can be successfully recovered by the co-overexpression approach. For instance, a deletion slightly larger than β′ΔSI3 (Δ932–1137 *versus* Δ943–1130 in β′ΔSI3) failed to assemble and yielded only insoluble β′Δ(932–1137) aggregates *in vivo*.

Role of β′SI3 in RNAP Function—Although both SI1 and SI2 in β exhibited some effects on the biochemical properties of RNAP (and were essential for full viability *in vivo*), the most dramatic effects on RNAP properties were observed for β′SI3. β′SI3 profoundly affected initiation and pausing by *E. coli* RNAP, in both cases accounting for many but not all of its differences from *B. subtilis* RNAP. We will consider the effects on initiation and pausing separately, although the underlying changes in RNAP's properties could be related.

β′SI3 stabilized open complexes formed at the T7 A1 promoter against disruption by heparin. This stabilization could reflect either direct contact with the downstream DNA, as previously proposed (16), or it could reflect indirect stabilization of the open complex if folding of β′SI3 is coupled to structural transitions in RNAP that occur upon open complex formation. Although direct DNA interaction is an attractive hypothesis, no cross-linking studies to date unambiguously document interaction of β′SI3 with the downstream DNA in open complexes. In the recently published crystal structure of *T. thermophilis* RNAP, the loop in β′G to which β′SI3 connects is positioned near the downstream DNA channel (11), also favoring its possible interaction with downstream DNA. However, this structure is of an enzyme that lacks β′SI3 and was obtained in the absence of DNA. The β′G loop is poorly ordered in both the yeast RNAPII and *T. aquaticus* RNAP crystal structures and was found in different conformations in the ordered portions of the two structures, suggesting that it is capable of significant motion. Further, β′SI3 itself is not visible in a recent electron microscopy crystal structure of *E. coli* RNAP, suggesting that it is even more mobile than the β′G loop (10). Thus, β′SI3 folding could be coupled either directly or indirectly to interactions of RNAP with downstream DNA. Many DNA-binding proteins undergo folding transitions in protein segments distant from the actual DNA binding surface (45). These indirect but coupled protein folding events are thought to contribute significantly to the avidity of protein-DNA interactions. Thus, stabilization of open complexes by β′SI3 could arise not via its direct interaction with the downstream DNA but via an indirect coupling of its folding to open complex formation. Further study will be required to distinguish these possibilities.

Both the direct and indirect explanation of β′SI3 function also could apply to its effect on transcriptional pausing. Both formation of paused transcription elongation complexes and

their slow escape are enhanced by an interaction of downstream DNA with RNAP (46–49). Thus, β′SI3 could strengthen the pause-enhancing downstream DNA interaction either directly or by the indirect mechanism described above. Interestingly, the effects on pausing of deleting β′SI3 are quite similar to the effects of deleting the β′ jaw domain of *E. coli* RNAP, for which a variety of data suggest a direct DNA interaction (49).

In the case of pausing, however, a third explanation is possible. β′G, in which β′SI3 is inserted, appears to cross-link to the RNA 3′ nt in the *his* paused transcription complex[3] as well as to the RNA 3′ nt in arrested complexes (50) and in certain artificial transcription complexes that may mimic paused transcription complexes (51). Further, amino acid substitutions in β′G, near the site of β′SI3 insertion, strongly affect chain elongation and transcriptional pausing and termination (40, 51). If movements of the β′G loop are involved in transcriptional pausing, it could readily explain why deletion of β′SI3 has such a strong effect on pausing. This also might explain why a partial deletion in β′SI3 (Δ1091–1130) as well as monoclonal antibody binding to β′1091–1130 in wild type RNAP greatly increase pausing (14), whereas complete deletion of β′SI3 (Δ943–1130) decreases pausing (Figs. 4 and 5). By altering the structure of β′SI3, both β′Δ(1091–1130) and antibody binding could interfere with movements of the β′G loop necessary for rapid nucleotide addition, whereas removal of β′SI3 could favor such movements.

Roles of βSI1 and βSI2 in Abortive Initiation—βSI1 and βSI2 both exhibited effects on initiation complexes, either in open complex longevity, abortive initiation, or both. Interestingly, both SIs increased abortive initiation significantly, although they decreased open complex longevity modestly (βSI1) or had no effect on it (βSI2). This suggests that RNAP must be exceptionally sensitive to changes in its structure during abortive initiation. This could be related to the sequential rearrangements of contacts between core RNAP and σ that are thought to occur during the initial stages of RNA synthesis (52). Perhaps βSI1 and βSI2 somehow facilitate the core-σ rearrangement during initial transcript synthesis. Alternatively, they could indirectly affect the conformation of the main channel of RNAP such that their removal increases the rate of abortive transcript release.

Sequence Insertions Are Ubiquitous in Bacterial RNAPs—Finally, we note that sequence insertions are not limited to those found in the RNAPs from enteric and thermophilic bacteria. A search of sequences now available for β and β′ subunits in bacteria revealed that sequence insertions are ubiquitous and that the positions of these insertions are not restricted to those observed in the enteric and thermophilic bacteria (Figs. 1 and 6). Rather, the sequence insertions differ in size, sequence, and location but are predicted to be surface-exposed (Fig. 6). Such a distribution of sequence insertions is consistent with the idea that they confer species-specific properties on core RNAP, either by modulating its enzymatic activity directly or through interactions with transcription factors. The relatively low degree of divergence among SIs in closely related proteobacterial species (see Introduction) also lends support to this idea.

This pattern of surface-exposed sequence insertions is reminiscent of the small subunits of eukaryotic RNAPs, which also are surface-exposed (at least in RNAPII); vary in composition to some extent among RNAPI, RNAPII, and RNAPIII; and can affect the biochemical properties of the enzyme yet are dispensable for core function. The RNAPII subunits 4 and 7, for instance, are readily dissociated from RNAP and affect its biochemical properties but are dispensable for nucleotide addition (53, 54); deletion of RNAPII subunit 9, like deletions in β′SI3, affects responses of the RNAPs to their respective RNA cleavage factors, GreB (14) and TFIIS (55). Perhaps the bacterial sequence insertions are like these small eukaryotic RNAP subunits, conferring distinctive properties on the RNAPs of different bacterial species much as the small RNAP subunits may confer distinctive properties on RNAPI, RNAPII, and RNAPIII.

Acknowledgments—We thank Seth Darst for help with molecular modeling and Josefine Ederth, Kati Geszvain, Rachel Mooney, Ruth Saecker, and Kesha Toulokhonov for helpful comments on the manuscript.

[3] K. Toulokhonov and R. Landick, unpublished results.

REFERENCES

1. Sweetser, D., Nonet, M., and Young, R. A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 1192–1196
2. Jokerst, R. S., Weeks, J. R., Zehring, W. A., and Greenleaf, A. L. (1989) *Mol. Gen. Genet.* **215,** 266–275
3. Zhang, G., Campbell, E. A., Minakhin, L., C., R., Severinov, K., and Darst, S. A. (1999) *Cell* **98,** 811–824
4. Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001) *Science* **292,** 1863–1876
5. Ebright, R. H. (2000) *J. Mol. Biol.* **304,** 687–698
6. Darst, S. A. (2001) *Curr. Opin. Struct. Biol.* **11,** 155–162
7. Severinov, K. (2000) *Curr. Opin. Microbiol.* **3,** 118–125
8. Sousa, R. (1996) *Trends Biochem. Sci.* **21,** 186–190
9. Opalka, N., Mooney, R. A., Richter, C., Severinov, K., Landick, R., and Darst, S. A. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 617–622
10. Darst, S. A., Opalka, N., Chacon, P., Polyakov, A., Richter, C., Zhang, G., and Wriggers, W. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 4296–4301
11. Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S., and Yokoyama, S. (2002) *Nature* **417,** 712–719
12. Nene, V., and Glass, R. E. (1984) *Mol. Gen. Genet.* **196,** 64–67
13. Borukhov, S., Severinov, K., Kashlev, M., Lebedev, A., Bass, I., Rowland, G. C., Lim, P. P., Glass, R. E., Nikiforov, V., and Goldfarb, A. (1991) *J. Biol. Chem.* **266,** 23921–23926
14. Zakharova, N., Bass, I., Arsenieva, E., Nikiforov, V., and Severinov, K. (1998) *J. Biol. Chem.* **273,** 24912–24920
15. Severinov, K., Kashlev, M., Severinova, E., Bass, I., McWilliams, K., Kutter, E., Nikiforov, V., Snyder, L., and Goldfarb, A. (1994) *J. Biol. Chem.* **269,** 14254–14259
16. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. A. (2000) *Science* **289,** 619–625
17. Artsimovitch, I., Svetlov, V., Anthony, L., Burgess, R. R., and Landick, R. (2000) *J. Bacteriol.* **182,** 6027–6035
18. Zillig, W. (1975) in *Biochemistry of the Cell Nucleus* (Hidvegi, E., Hidvegi, E. J., Sumegi, J., and Venetianer, P., eds) Vol. 33, pp. 239–251, Amsterdam North-Holland
19. Tang, H., Severinov, K., Goldfarb, A., and Ebright, R. H. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 4902–4906
20. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol.* **185,** 60–89
21. Anthony, L. C., Artsimovitch, I., Svetlov, V., Landick, R., and Burgess, R. R. (2000) *Protein Expression Purif.* **19,** 350–354
22. Gribskov, M., and Burgess, R. R. (1983) *Gene (Amst.)* **26,** 109–118
23. Artsimovitch, I., and Landick, R. (1998) *Genes Dev.* **12,** 3110–3122
24. Feng, G., Lee, D. N., Wang, D., Chan, C. L., and Landick, R. (1994) *J. Biol. Chem.* **269,** 22282–22294
25. Landick, R., Wang, D., and Chan, C. (1996) *Methods Enzymol.* **274,** 334–352
26. Fukuda, R., and Ishihama, A. (1974) *J. Mol. Biol.* **87,** 523–540
27. Hager, D. A., Jin, D. J., and Burgess, R. R. (1990) *Biochemistry* **29,** 7890–7894
28. Hawley, D. K., Johnson, A. D., and McClure, W. R. (1985) *J. Biol. Chem.* **260,** 8618–8626
29. Buc, H., and McClure, W. R. (1985) *Biochemistry* **24,** 2712–2723
30. Craig, M. L., Tsodikov, O. V., McQuade, K. L., Schlax, P. E., Jr., Capp, M. W., Saecker, R. M., and Record, M. T., Jr. (1998) *J. Mol. Biol.* **283,** 741–756
31. Saecker, R. M., Tsodikov, O. V., McQuade, K. L., Schlax, P. E., Capp, M. W., and Record, M. T. (2002) *J. Mol. Biol.* **319,** 649–671
32. Calles, B., Monsalve, M., Rojo, F., and Salas, M. (2001) *J. Mol. Biol.* **307,** 487–497
33. Rojo, F., Nuez, B., Mencia, M., and Salas, M. (1993) *Nucleic Acids Res.* **21,** 935–940
34. Whipple, F. W., and Sonenshein, A. L. (1992) *J. Mol. Biol.* **223,** 399–414
35. Nechaev, S., Chlenov, M., and Severinov, K. (2000) *J. Biol. Chem.* **275,** 25516–25522
36. Pfeffer, S. R., Stahl, S. J., and Chamberlin, M. J. (1977) *J. Biol. Chem.* **252,** 5403–5407
37. Toulokhonov, I., Artsimovitch, I., and Landick, R. (2001) *Science* **292,** 730–733
38. Minakhin, L., Bhagat, S., Brunning, A., Campbell, E. A., Darst, S. A., Ebright, R. H., and Severinov, K. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 892–897
39. Severinov, K., Mustaev, A., Kashlev, M., Borukhov, S., Nikiforov, V., and Goldfarb, A. (1992) *J. Biol. Chem.* **267,** 12813–12819
40. Weilbaecher, R., Hebron, C., Feng, G., and Landick, R. (1994) *Genes Dev.* **8,** 2913–2917
41. Palm, P., Heil, A., Boyd, D., Grampp, B., and Zillig, W. (1975) *Eur. J. Biochem.*

53, 283–291
42. Zalenskaya, K., Lee, J., Gujuluva, C. N., Shin, Y. K., Slutsky, M., and Goldfarb, A. (1990) *Gene* (*Amst.*) **89,** 7–12
43. Borukhov, S., and Goldfarb, A. (1993) *Protein Expression Purif.* **4,** 503–511
44. Kashlev, M., Nudler, E., Severinov, K., Borukhov, S., Komissarova, N., and Goldfarb, A. (1996) *Methods Enzymol.* **274,** 326–334
45. Spolar, R. S., and Record, M. T., Jr. (1994) *Science* **263,** 777–784
46. LaFlamme, S. E., Kramer, F. R., and Mills, D. R. (1985) *Nucleic Acids Res.* **13,** 8425–8440
47. Lee, D. N., Phung, L., Stewart, J., and Landick, R. (1990) *J. Biol. Chem.* **265,** 15145–15153
48. Chan, C., Wang, D., and Landick, R. (1997) *J. Mol. Biol.* **268,** 54–68
49. Ederth, J., Artsimovitch, I., Isaksson, L., and Landick, R. (2002) *J. Biol. Chem.* **277,** 37456–37463
50. Markovtsov, V., Mustaev, A., and Goldfarb, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 3221–3226
51. Epshtein, V., Mustaev, A., Markovtsov, V., Bereshchenko, O., Nikiforov, V., and Goldfarb, A. (2002) *Mol. Cell* **10,** 623–634
52. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., and Darst, S. A. (2002) *Science* **296,** 1285–1290
53. Edwards, A. M., Kane, C. M., Young, R. A., and Kornberg, R. D. (1991) *J. Biol. Chem.* **266,** 71–75
54. Orlicky, S. M., Tran, P. T., Sayre, M. H., and Edwards, A. M. (2001) *J. Biol. Chem.* **276,** 10097–10102
55. Awrey, D. E., Weilbaecher, R. G., Hemming, S. A., Orlicky, S. M., Kane, C. M., and Edwards, A. M. (1997) *J. Biol. Chem.* **272,** 14747–14754
56. Dubendorff, J. W., and Studier, F. W. (1990) *J. Mol. Biol.* **219,** 45–59
57. Landick, R., Stewart, J., and Lee, D. (1990) *Genes Dev.* **4,** 1623–1636
58. Severinov, K., and Darst, S. A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 13481–13486
59. Wang, D., Meier, T., Chan, C., Feng, G., Lee, D., and Landick, R. (1995) *Cell* **81,** 341–350
60. Artsimovitch, I., and Landick, R. (2002) *Cell* **109,** 193–203
61. Fiil, N. P., Bendiak, D., Collins, J., and Friesen, J. D. (1979) *Mol. Gen. Genet.* **173,** 39–50
62. Yuzenkova, J., Delgado, M., Nechaev, S., Savalia, D., Epshtein, V., Artsimovitch, I., Mooney, R., Landick, R., Farrias, R. N., Salomon, R., and Severinov, K. (2002) *J. Biol. Chem.* **277,** 50857–50875

Co-overexpression of *Escherichia coli*RNA Polymerase Subunits Allows Isolation and Analysis of Mutant Enzymes Lacking Lineage-specific Sequence Insertions

Irina Artsimovitch, Vladimir Svetlov, Katsuhiko S. Murakami and Robert Landick

J. Biol. Chem. 2003, 278:12344-12355.
doi: 10.1074/jbc.M211214200 originally published online January 2, 2003
