INTRODUCING ARCHAEBACTERIAL AND CHLOROPLAST SPLIT SITES IN THE β AND β' SUBUNITS OF ESCHERICHIA COLI RNA POLYMERASE

(Received for publication, April 27, 1996, and in revised form, July 24, 1996)

Konstantin Severinov‡§, Arkady Mustaev¶, Aleksandr Kukarin†, Oriana Muzzin‡, Irina Bass‡, Seth A. Darst‡, and Alex Goldfarb¶

From ‡The Rockefeller University, New York, New York 10021, the ¶Public Health Research Institute, New York, New York 10016, and the Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, Russia 123182

The β and β' subunits of Escherichia coli DNA-dependent RNA polymerase are highly conserved throughout euarchaeal and eukaryotic kingdoms. However, in some archaebacteria and chloroplasts, the corresponding sequences are “split” into smaller polypeptides that are encoded by separate genes. To test if such split sites can be accommodated into E. coli RNA polymerase, subunit fragments encoded by the segments of E. coli rpoB and rpoC genes corresponding to archaebacterial and chloroplast split subunits were individually expressed in E. coli. The purified fragments, when mixed in vitro with complementing intact RNA polymerase subunits, yielded an active enzyme capable of catalyzing the phosphodiester bond formation. Thus, the large subunits of eubacteria and eukaryotes are composed of independent structural modules corresponding to the smaller subunits of archaebacteria and chloroplasts.

DNA-dependent RNA polymerase (RNAP)1 is the central enzyme of gene expression and a major target for genetic regulation. The structure of cellular RNAP has been highly conserved in evolution (1–4), as documented by the high level of similarity between the primary sequences of RNAP subunits from bacteria to man (5–8). New models of RNAP function postulate a substantial degree of intramolecular motion within the transcription complex during transcription (9–11). Movement of enzyme parts relative to each other could be realized by structurally independent modules that are connected by flexible loops or hinges. Identification of such modules would afford new insights into the mechanism of RNAP function.

Modular organization of RNAP is also suggested by sequence comparisons between the β and β' subunit homologs from eubacteria, archaebacteria, and eukaryotes. Together, the two subunits display 17 colinear segments of high sequence similarity separated by regions that are poorly conserved (5–8). In some organisms, the non-conserved regions contain large gaps or insertions compared with the Escherichia coli β and β' subunits. Large deletions in two non-conserved regions of β do not affect RNAP assembly and basic function in vitro (12, 13). These regions, centered around residues 300 and 1000, have been termed dispensable regions I and II, respectively. Together, the two dispensable regions comprise more than 25% of the total length of β. It has been demonstrated that physical continuity of the β subunit in dispensable regions is not necessary for RNAP assembly and basic function in vitro (14, 15). These results indicate that β is comprised of the N-terminal (amino acids 1–235), middle (amino acids 235–950), and C-terminal (amino acids 950–1342) domains that need not be covalently connected for the enzyme to assemble and function in vitro.

The β' subunit homolog from archaebacterial genus Sulfolobus is encoded by two separate genes (16). The split site corresponds to codon 850 of the rpoC gene, which encodes the β' subunit in E. coli. RNA polymerase gene organization of halo-phylic and methanobacteria reveals even higher levels of complexity since both their rpoB (the gene encoding β) and rpoC homologs are split (16–18). The split site in the β homolog projects close to codon 650 of E. coli rpoB, whereas the β' split occurs at the same site as in Sulfolobus (16). The rpoC genes of cyanobacteria and some chloroplasts are split around position 550 of the E. coli sequence (19, 20).

RNA polymerases purified from these organisms and organelles have subunit composition reflecting the genetic splits (21–23). At least in some cases transcription activity of such naturally split enzymes has been demonstrated in vitro (22–25).

In this report, we demonstrate that all of the known β and β' split sites found in natural RNAPs can be reconstructed in E. coli RNAP without the loss of in vitro function.

MATERIALS AND METHODS

Genetic Techniques—The pUC19-based pMSk250 rpoB expression plasmid was used to generate the archaebacterial split site in the rpoB gene. The plasmid is a derivative of pMKs250 (26) and has a unique HindIII site introduced at codon 643 of the wild-type rpoB gene. The HindIII site was introduced using the site-directed mutagenesis procedure described previously (27). The sequence of the β subunit expressed from pMSk250 remains wild type due to degeneracy of the genetic code. Deletion mutations were generated by linearizing pMSk250 with HindIII, treating it with Bal31 exonuclease for a limited time, flushing the ends with Klenow enzyme, recircularizing the plasmid, treating it with HindIII, and transforming it into XL1-blue host strain. Recombinant clones overproducing the “full-sized” β polypeptide upon addition of isopropyl-1-thio-β-D-galactopyranoside were selected by SDA-polyacrylamide gel electrophoresis on 7.5% Phast gels (Pharmacia Biotech Inc.). The in vivo function of the mutant rpoB genes was assayed using the E. coli AlJ6207 suppressor strain as described previously (26).

The plasmid overproducing the N-terminal fragment of β was constructed by linearizing pMSk250 with HindIII, flushing the ends with Klenow enzyme, and religating. This procedure destroys the rpoB reading frame after codon 643 and generates a β polypeptide amber frag-
ment terminated shortly after the frameshift site. The plasmid expressing $\beta'$-lacZ was obtained by retaining the HindIII-Spl fragment of pMKSe250 into appropriately treated pUC19. The resulting plasmid expresses a translational fusion of the first five amino acids of the $\beta$-galactosidase $\alpha$ peptide followed by $\beta'$ amino acids 643-1342.

Plasmids expressing N-terminal fragments of $\beta'$ were generated from pMKa201 (28) by cutting with PstI (cuts twice at rpoC codons 577 and 593) or SauI (cuts once at codon 577), flushing the ends with Klenow enzyme (inhibits off-line digestion). The plasmid expressing a fusion of the first 8 amino acids of the $\beta$-galactosidase $\alpha$ peptide followed by the $\beta'$ sequence. E. coli XL1-blue was used as the expression host for all the plasmids described above.

The plasmid expressing $\beta'$(820–1407) was obtained by treating pMKa201 with BspHI (cuts once at rpoC codon 820) and ScaI (cuts in the polynucleoty upstream of the structural gene and at codon 544 of the structural gene) and recircularizing. The resulting plasmid expresses a translational fusion of the first 8 amino acids of the $\beta$-galactosidase $\alpha$ peptide followed by the $\beta'$ sequence. E. coli BL21(DE3) cells were used to express this fragment.

All of the overexpressed RNAP subunit fragments were found in inclusion bodies when induced with 1 mM isopropyl-1-thio-$\beta$-galactopyranoside at 37 °C.

**RNAP Reconstitution**—RNAP was reconstituted according to previously published procedures (29, 30). The molar ratio of $\alpha$ and intact $\beta$ or $\beta'$ in the recombinant RNAP preparations was 1:4, which is reflected by this ratio of the intact $\beta$ and $\beta'$ subunit fragments was added to 8-fold molar excess relative to $\alpha$. After reconstitution and thermoactivation in the presence of $\alpha^9$, RNAP preparations were either directly used for affinity labeling or further purified by fast protein liquid chromatography/gel filtration on a Superose-6 column (Pharmacia) as described (29), concentrated by filtration through a C-100 concentrator (Amicon, Inc.) to ~1 mg/ml, and stored in 50% glycerol storage buffer at −20 °C.

**Affinity Reagents, Affinity Labeling, and Cross-linked Product Analysis**—Synthesis of the aldehyde derivative of AMP is described elsewhere (31). For Lys$^{[13]}$O$\beta$-specific affinity labeling, 20 $\mu$l of RNAP reconstitution mixture (5 $\mu$g of RNAP holoenzyme in a buffer containing 50 mM Tris-HCl, pH 7.9, 200 mM KCl, 10 mM MgCl$_2$, 10 mM MnCl$_2$, 10% glycerol, and 5 mM $\beta$-mercaptoethanol) was supplemented with the initiating substrate analog (0.5–1.0 mM), and reactions were incubated for 15 min at 37 °C. 100 ng of a 135-base pair T7 A1 promoter-containing DNA fragment (32) and sodium borohydride (10 mM) were added, and incubation at 37 °C was continued for 10 min. Reactions were transferred to an ice bath and then incubated for 30 s in the presence of 0.3 $\mu$m $[\alpha-\beta^{32}P]$UTP (3000 Ci/mm). Reactions were terminated by adding an equal volume of SDS-containing Laemmli loading buffer, and proteins were resolved by SDS-polyacrylamide gel electrophoresis in precast 8% Tris/glycine gels (Novex). Gels were stained with Coomassie Blue, and affinity-labeling products were visualized by autoradiography. Control experiments confirmed that in accordance with the published data (31) $^{32}$P-labeling of the $\beta$ subunit strictly depended on the addition of template DNA and the cross-linking reagent. To affinity label the $\beta'$ subunit, we used a novel reagent, 3'-deoxy-3'-bromacetamido-TTP (see Fig. 1). The synthesis of the reagent was performed essentially following the procedure for a similarly derivatized ATP (33) and will be described elsewhere. Immobilized RNAP was used to form open complexes on a T7 A2 promoter-containing DNA fragment (34) (initial transcribed sequence GCU). Transcription was initiated in 20 $\mu$l of reconstitution mixture (see above) from a Rif(G) chimeric molecule (35) by addition of 0.3 mM $^{[\alpha-\beta^{32}P]}$CTP (3000 Ci/mm), specified by the +2 position of the promoter. Immobilized transcription reactions were washed once with 1.5 ml of buffer containing 40 mM Tris-HCl, 100 mM KCl, 2.5 mM MgCl$_2$, pH 7.9. 3'-deoxy-3'-bromacete- 

In **In Vitro Transcription**—The 135-base pair DNA fragment with T7A1 promoter subunit was used for transcription experiments. Transcription reactions were performed essentially as described (27). The 20-$\mu$l reactions contained 40 mM Tris-HCl, pH 7.9, 100 mM KCl, 10 mM MgCl$_2$, 1 mM rpoN, 10 pmol of template DNA, 1 mM Cpa, 100 $\mu$M CTP and GTP, 50 $\mu$M ATP, and 10 $\mu$l $[\alpha-\beta^{32}P]$UTP (300 Ci/mm). Reactions proceeded for 15 min at 37 °C and were terminated by the addition of 20 $\mu$l of 6 M urea-containing loading buffer. Reaction products were resolved by electrophoresis in 23% (20:3) polyacrylamide gel in the presence of 6 $\mu$l urea and visualized by autoradiography.

**RESULTS**

**Generation of $\beta$ and $\beta'$ Subunit Fragments and RNAP Reconstitution**—Plasmids expressing the E. coli $\beta$ and $\beta'$ fragments corresponding to the naturally occurring split sites discussed in the introduction were generated as described under “Materials and Methods.” All of the subunit fragments were overexpressed and recovered from inclusion bodies. Subunit fragments were solubilized in buffer containing 6 M guanidine-HCl, complementing subunit fragments together with the missing intact RNAP subunits were added, and the denaturant was dialyzed away at conditions favoring RNAP assembly from intact $\alpha$, $\beta$, and $\beta'$ subunits (29, 30).

**Strategy for in Vitro Functional Analysis**—The function of RNAPs with split subunits was tested using the selective affinity labeling technique (31). The technique employs substrate NTP analogs that can be incorporated into nascent RNA and then cross-linked to either $\beta$ or $\beta'$. The functional capacity of RNAPs reconstituted from split subunits to form a phosphodiester bond is manifested by radioactive labeling of subunit fragments containing the cross-link site. This test was not compromised by contamination of the preparations with WT RNAP, which is revealed, if present, by labeled full-sized $\beta$ and/or $\beta'$.

The function of reconstituted RNAPs with split $\beta$ was tested using promoter-dependent affinity labeling with a novel cross-linkable reagent positioned at the 3'-end of the nascent RNA in the ternary transcription complex (Fig. 1, Reagent B). The reagent was used in combination with a chimeric Rif(G) initiator molecule, fully analogous to the recently described Rif(A) compound (35). Rif(G) is stably bound to RNAP in the open complex through its Rif moiety, allowing us to prepare transcription complexes containing RNA only 2–3 nucleotides in length.

At the first stage of the reaction, an open complex between RNAP and a T7 A2 promoter-containing DNA fragment was formed in the presence of Rif(G). The initial transcribed sequence of this promoter is GCT. At the second stage of the reaction, bound Rif(G) was extended with $[\alpha-\beta^{32}P]$CTP, specified by the +2 position of the promoter. At the third stage of the reaction, the cross-linkable analog of dTTP, 3'-deoxy-3'-bromacetamido-TTP, specified by the +3 position of the promoter was added. To allow incorporation of the dTMP analog instead of UMP, the reaction buffer of the third stage contained Mn$^{2+}$ instead of Mg$^{2+}$. Cross-linking occurred spontaneously and resulted in attachment of radioactive Rif-GCU to the $\beta'$ subunit conserved segment G in the C-terminal portion of $\beta'$ (amino acids 912–939). Cross-linking to an uncharacterized site in $\beta'$ also occurred. The levels of cross-linking to $\beta$ constituted less than 10% of the $\beta'$ cross-linking (Fig. 3B, lane 5).

**Split $\beta'$ Subunit Assemblies in Vitro into Catalytically Proficient RNA Polymerase**—Results of highly selective affinity labeling in reconstitution mixtures containing $\beta'$ subunit split at the chloroplast site are presented in Fig. 2A. In this experiment, the C-terminal fragment of the $\beta'$ subunit from amino
acid 544 to 1407 was reconstituted with the N-terminal β′-(1–577) and intact β, α, and σ. The two β′ fragments correspond to the γ and the β′ subunits of RNAP from cyanobacteria and chloroplasts. Reconstitution reaction containing intact β′, β, α, and σ was used as a positive control (lanes 1 and 3). After affinity labeling with the β′-specific Reagent A, reaction products were resolved by SDS-polyacrylamide gel electrophoresis. The gel was stained with Coomassie Blue (left panel), and labeling products were revealed by autoradiography (right panel). Radioactive labeling of the larger, C-terminal fragment of β′ (lane 4, β′c) verifies that the split β′ assembled into functional enzyme since labeling requires the formation of two phosphodiester bonds. Control experiments established that labeling was strictly promoter-dependent; no labeling was observed if only one β′ fragment was used in the reconstitution reaction (data not shown).

Similarly, affinity labeling of reconstitution mixtures containing the C-terminal fragment of the β′ subunit from amino acid 820 to 1407, the N-terminal β′-(1–877), and the intact β, α, and σ subunits resulted in the labeling of the smaller fragment β′-(820–1407), corresponding to the C component of RNAP from archae (Fig. 2B, lane 4, β′c). Thus, the E. coli RNAP incorporating the archaebacterial split in β′ is also able to catalyze formation of the first two phosphodiester bonds.

The above experiments demonstrate that split β′-RNAPs are able to bind promoter DNA and catalyze formation of the first two phosphodiester bonds. Affinity labeling experiments also demonstrate that the split enzymes are essentially free of contaminating WT RNAP. To establish if the split β′ enzymes are able to synthesize longer RNA, transcription of a DNA fragment containing the A1 promoter of bacteriophage T7 was performed (Fig. 2C). The results demonstrate that the enzyme with the archaël split produced essentially the wild-type pattern of transcription products. However, the chloroplast split caused a severe defect in promoter clearance (lane 2). The split enzyme grossly overproduced the first abortive product, CpaP, at the expense of the longer products. At longer exposures, longer RNA products synthesized by the split enzyme became visible. The pattern of the longer products was qualitatively different from the wild-type pattern and hence was due to activity of the split enzyme rather than contaminating WT RNAP (data not shown).

**Splitting the β′ Subunit at Archaebacterial Site**—As mentioned in the Introduction, RNAP from halophylic archaebacteria have both their β and β′ homologs split. To investigate if these split sites can be incorporated in the E. coli system, we first wanted to demonstrate that β split in the archaebacterial site assembles into functional RNAP (Fig. 3A). The function of the split enzyme was readily manifested by affinity labeling of the C-terminal fragment β′-(843–1342), which contains Lys71065.
with Reagent B (note that the two fragments corresponding to archael $b$ and $b^*$ components have the same mobility on the gel). The split enzyme was also able to synthesize longer RNA indistinguishably from the WT RNAP (Fig. 3C).

The archael split site occurred in the evolutionarily variable region of $b$, between conserved regions D and E (Fig. 4). The two previously characterized split sites in $b$ also were introduced in highly variable and functionally dispensable regions of the polypeptide. To further probe the functional role of this region of $b$, three short deletions were generated (Fig. 4A). Mutant $rpoB$ genes, $\Delta(644-646)$, $\Delta(641-656)$, and $\Delta(627-649)$, could be the only source of $b$ in the cells as judged by the ability of plasmids expressing the mutant genes to complement an amber mutation in the chromosomal copy of $rpoB$ in the E. coli AJ6207 strain (26). Thus, the archael split occurred in the evolutionarily variable and functionally dispensable region of $b$.

$\beta$ and $\beta'$ Subunits Split Simultaneously at Archaeobacterial Sites Assemble in Vitro into a Functional Enzyme—Having established that the archael $\beta$ split site can be incorporated in functional E. coli RNAP, we asked if the archael $\beta$ and $\beta'$ split sites can be combined in a single E. coli RNAP molecule (such an arrangement is found in RNA polymerases from halophilic archaea and methanobacteria). The split enzyme was

![Fig. 3. Splitting the $\beta$ subunit and $\beta$ and $\beta'$ simultaneously at archaeabacterial sites. Recombinant $\beta$ and $\beta'$ fragments, or $\beta$ and $\beta'$ fragments, were incubated with a complement of intact RNAP subunits at conditions favoring RNAP reconstitution. After affinity labeling, reaction products were analyzed as described for Fig. 2. A, splitting $\beta$ in the archaeabacterial split site. N-terminal fragment $\beta$(1–643) and C-terminal fragment $\beta$(643–1342) were mixed in reconstitution reactions. Affinity labeling was performed with Reagent B. B, splitting $\beta$ and $\beta'$ simultaneously in archaeabacterial sites. $\beta$(1–643) and $\beta'$(643–1342) were mixed with $\beta^*$ (641–877) and $\beta^*$ (627–1407) in reconstitution reactions. Reconstitution mixtures were modified with Reagent B (lanes 3 and 4) or A (lane 5). C, transcription by split RNAPs. Autorad., autoradiograph; holo, holoenzyme.]

![Fig. 4. Domain organization of RNA polymerase $\beta$ subunit and deletion mutagenesis around the archaeabacterial split site. The bar represents the 1342-amino acid-long $\beta$ subunit of E. coli. The lettered gray boxes designate evolutionarily conserved segments. Open boxes represent large deletions found in $\beta$ homologs from Gram-positive bacteria and chloroplasts. Dispensable regions are represented by functional deletions (black lines, in scale). The sites of known rifampicin- ($\text{Rif}^R$) and streptolidygin-resistant ($\text{Stl}^R$) mutations (27, 47, 48), as well as the sites of cross-links to the phosphates of the initiating nucleotide, are also indicated (49, 50). The area between conserved segments D and E is enlarged at the bottom. The sequence from E. coli (E. c., top row) (51) is compared with the corresponding sequences from Pseudomonas putida (P. p.) (52), Bacillus subtilis (B. s.) (53), Thermatoga maritima (T. m.) (54), and chloroplasts from tobacco (T.) (55). The deletions obtained in this work are shown at the top of the E. coli sequence. The sites of the “archael” split obtained in this work and functional splits in dispensable regions (DRI and DRII) (15) are shown by black vertical arrows.]
reconstituted in vitro with intact α and σ and assayed in the affinity labeling assay. In separate reactions, labeling with both β and β′-specific reagents was used to demonstrate that the double split RNAP is essentially free of contamination with the functional WT as well as single-split enzymes. The results of the experiment presented on Fig. 3B demonstrate that E. coli RNAP harboring both of the archaeabacterial split sites was labeled at the C-terminal fragment of β by Lys1065-specific reagent (lane 4). The double split enzyme was also labeled at the C-terminal fragment of β′ and, with lesser efficiency, at one of the β fragments that are not resolved at the gel (lane 5). No labeling of the full-sized β and/or β′ was apparent. We conclude that our preparation of the double split enzyme is functional and essentially free of contamination. The double split enzyme was also able to synthesize longer RNA, and the pattern of transcription products was only slightly changed when compared with the wild-type RNAP pattern (Fig. 3C).

DISCUSSION

Modular Organization of β and β′—The principal conclusion of this work is that the segments of the E. coli β and β′ subunits, which are demarcated by split sites in archaeabacteria and chloroplasts, need not be covalently linked to assemble into active RNAP capable of locating the promoter, catalyzing phosphodiester bond formation, and, with the exception of the chloroplast split site in β′, elongating the RNA transcript in vitro. As is argued elsewhere (15), split sites that allow assembly into functional enzyme probably define rough boundaries of structural domains. Therefore, the data reported here, together with our previous work (15), argue that the β subunit is comprised of at least four distinct modules consisting of amino acids 1–235, 235–650, 650–950, and 950–1342 (Fig. 4).

Similarly, our results establish that β′ is comprised of at least three independent modules: N-terminal (amino acids 1–550), middle (amino acids 550–850), and C-terminal (amino acids 850–1407) (Fig. 5). The actual domain organization of β′ is almost certainly more complex. For example, phylogenetic analysis reveals that ~200 amino acids between conserved regions G and H of E. coli are absent from the β′ homologs in Gram-positive bacteria (Fig. 5) (36, 37). In chloroplasts, the same region harbors insertions of up to 400 amino acids (38, 39). These data make it almost certain that the C-terminal module of β′ defined in this work actually consists of at least two domains, separating regions G and H.

Functional Implications of β and β′ Modules—Each of the large subunit modules contains evolutionarily conserved segments (Figs. 4 and 5). The conservation of sequence is believed to reflect functional significance. Thus, the structural modules defined in this work may represent distinct functional entities of RNAP. New models of the transcription mechanism postulate substantial conformational transitions in the RNAP molecule during promoter clearance, elongation, pausing, and termination (11, 40–44). It is conceivable that movements of the modules relative to each other constitute the basis of these transitions.

If the internal shifting of the modules plays a functional role in RNAP movement, the split enzyme could demonstrate elongation defects. Indeed, some of the previously described split β enzymes displayed defects in elongation (15). In this work, we demonstrate a dramatic effect of the chloroplast split in β′ on promoter clearance. The exact nature of this defect is currently being investigated, but it is tempting to speculate that the split enzyme falls apart as it stretches when attempting to clear the promoter. We note, however, that RNAP purified from chloroplasts and cyanobacteria is able to synthesize long RNA (23, 25). Compared with the E. coli enzyme, chloroplast RNAP has, besides the split β′, large insertions in the variable region of the β homolog as well as deletions in the variable regions of the β homolog (see Figs. 4 and 5). It is possible that these changes compensate for the functional defects caused by the split in the E. coli system.

Subunit Modules in RNAP Assembly—In E. coli, RNAP assembly proceeds in a defined pathway, α-α2-α2β-α2β′ (45). This pathway appears to have been conserved between E. coli and yeast (46). The assembly pathway has to be more complex in organisms with split β, β′ homologs. The ability to model these split sites in the well-studied E. coli system opens new approaches to the analysis of RNAP assembly and subunit interaction.

REFERENCES
1. Darst, S. A., Kubalek, R. W., and Kornberg, R. D. (1989) Nature 340, 730–732
2. Darst, S. A., Edwards, A. M., Kubalek, E. W., and Kornberg, R. D. (1991) Cell 66, 121–128
3. Schultz, P., Celia, H., Riva, M., Sentenac, A., and Oudet, P. (1993) EMBO J.
