Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of *E. coli* ribosomal RNA promoters

Jared T. Winkelman[a,1], Pete Chandrangsu[a,1], Wilma Ross[a], and Richard L. Gourse[a,2]

[a]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI 53706

Edited by Jeffrey W. Roberts, Cornell University, Ithaca, NY, and approved February 23, 2016 (received for review November 9, 2015)

Most *Escherichia coli* promoters initiate transcription with a purine 7 or 8 nt downstream from the –10 hexamer, but some promoters, including the ribosomal RNA promoter *rrnB* P1, start 9 nt from the –10 element. We identified promoter and RNA polymerase determinants of this noncanonical *rrnB* P1 start site using biochemical and genetic approaches including mutational analysis of the promoter, Fe^{2+} cleavage assays to monitor template strand positions near the active-site, and Bpa cross-linking to map the path of open complex DNA at amino acid and nucleotide resolution. We find that mutations in several promoter regions affect transcription start site (TSS) selection. In particular, we show that the absence of strong interactions between the discriminator region and σ region 1.2 and between the extended –10 element and σ region 3.0, identified previously as a determinant of proper regulation of rRNA promoters, is also required for the unusual TSS. We find that the DNA in the single-stranded transcription bubble of the *rrnB* P1 promoter complex expands and is "scrunched" into the active site channel of RNA polymerase, similar to the situation in initial transcribing complexes. However, in the *rrnB* P1 open complex, scrunching occurs before RNA synthesis begins. We find that the scrunched open complex exhibits reduced abortive product synthesis, suggesting that scrunching and unusual TSS selection contribute to the extraordinary transcriptional activity of rRNA promoters by increasing promoter escape, helping to offset the reduction in promoter activity that would result from the weak interactions with σ.

RNA polymerase | transcription start site selection | open complex scrunching | ribosomal RNA promoters | promoter escape

To initiate transcription, RNA polymerase (RNAP) and promoter DNA participate in a multistep binding reaction that results in unwinding of at least one turn of DNA and placement of the template strand into the active site. Once positioned, the initiating nucleoside triphosphate and the second NTP pair with the template, and the first phosphodiester bond is catalyzed.

There is considerable information about the structure of open complexes and the position of the transcription start site (TSS) for consensus promoters and engineered scaffolds. However, perfect consensus promoters are not actually found in wild-type *Escherichia coli* cells, and little is known about TSS selection on native promoters and about how variation in promoter sequence affects TSS position. Comparison of the TSS for natural and synthetic $E\sigma^{70}$-dependent promoters indicates that a purine 7–8 bp downstream from the last base in the –10 element is typically used for initiation, and in some cases, two or more adjacent positions in an individual promoter are used (1–3). A pyrimidine at nontemplate strand position –1 is preferred (Fig. 1*A*), because the corresponding purine on the template strand makes favorable stacking interactions with the initiating NTP (4). Transcripts initiating as far as 12 bp downstream from the –10 hexamer have been reported but are very uncommon (5, 6). Changes in nucleotide concentrations alter the TSS at some promoters, and the identity of the 5′ end of the transcript can affect transcription attenuation, reiterative transcription, translation efficiency, or RNA stability, regulating gene expression (7).

A recent single-molecule fluorescence resonance energy transfer study showed there can be heterogeneity in transcription bubble size in open complexes, even for an individual promoter (8). Thus, spontaneous bubble expansion and contraction can account for multiple start sites at some promoters. The variation in TSS from an individual promoter implies there is flexibility in the placement of template strand DNA by RNAP relative to the active site of the enzyme.

After formation of the first phosphodiester bond during transcription initiation, the next 5–12 nucleotide addition steps proceed through a scrunching mechanism in which the trailing edge of RNAP does not move with respect to DNA and 1 bp of downstream DNA is pulled into RNAP for every nucleotide added to the growing RNA chain (9, 10). Strain generated during initial transcription leads to formation of a high-energy state, and this energy can be released by either breaking promoter contacts to form an elongation complex or by releasing the nascent RNA as an abortive transcript. At most promoters, multiple rounds of synthesis and release of abortive transcripts take place before promoter contacts are broken, and this can be a rate-limiting step at some promoters (11).

Here we use the *E. coli* ribosomal RNA promoter *rrnB* P1 to investigate the promoter sequences and interactions with RNAP responsible for its unusual TSS (Fig. 1). RNAP initiates transcription from *rrnB* P1 in vitro and in vivo predominantly at an adenine 9 bp downstream from the end of the –10 hexamer rather than at the adenine 6 bp downstream from the –10 hexamer (12–14), the TSS position predicted from the TSSs at other

Significance

Although it has long been appreciated that there are multiple RNA polymerase recognition elements in promoters, the determinants of transcription start site (TSS) selection are poorly understood. We find here that the absence of strong interactions between σ and the –10 element region of *Escherichia coli* rRNA promoters results in expansion of the transcription bubble prior to nucleotide addition, a phenomenon referred to as scrunching. Open complex scrunching accounts for the unusually long distance between the –10 element and the TSS and appears to contribute to the extraordinary activity of these promoters by minimizing abortive product formation, thereby increasing transcriptional output. These data provide insights into the determinants of TSS selection, promoter function, and promoter evolution.

Author contributions: J.T.W., P.C., W.R., and R.L.G. designed research; J.T.W. and P.C. performed research; J.T.W., P.C., W.R., and R.L.G. analyzed data; and J.T.W., P.C., W.R., and R.L.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]J.T.W. and P.C. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: rgourse@bact.wisc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1522159113/-/DCSupplemental.

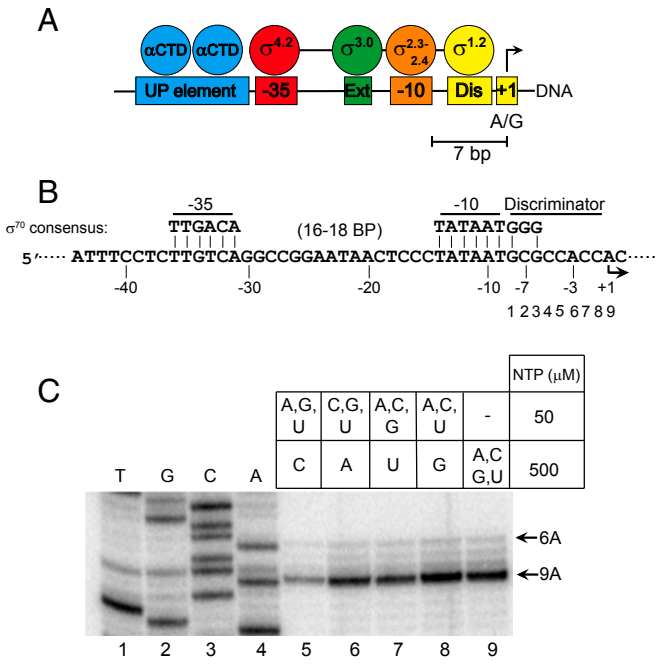

Fig. 1. Nucleotide concentration does not affect TSS selection of *rrnB* P1. (*A*) Sequence-specific contacts between RNAP ($E\sigma^{70}$) and promoter DNA (adapted from ref. 15). UP element–αCTD interaction, blue; –35 hexamer–σ region 4.2 interaction, red; extended –10 element–σ region 3.0 interaction, green; –10 hexamer–σ regions 2.3–2.4 interaction, orange; discriminator–σ region 1.2 interaction, yellow. Transcription typically initiates with a purine 7 bp downstream from the –10 element (1). (*B*) Comparison of the *rrnB* P1 promoter with consensus $E\sigma^{70}$ promoter elements. Matches to consensus hexamers are indicated. Promoter numbering is with respect to the wild-type TSS, designated as +1. Alternative numbering for start site region, 1–9 bp downstream of the –10 element, is indicated. (*C*) Start sites of *rrnB* P1 transcription determined by primer extension mapping of 5′-ends of transcripts generated in vitro at the indicated NTP concentrations. Extension products were separated on a denaturing acrylamide gel with a DNA sequence ladder generated using the same primer.

promoters (1). Using a library of 38 *rrnB* P1 promoter variants, we find that mutations at a variety of positions can alter TSS selection, implicating several specific RNAP–promoter interactions in the selection mechanism. Site-specific RNAP–DNA cross-linking indicates that the wild-type *rrnB* P1 TSS results from open complex DNA scrunching, in which downstream DNA is pulled into RNAP before nucleotide addition, allowing initiation of transcription farther downstream from the –10 element than is typical. Promoter mutations that shift the position of the TSS reduce open complex scrunching and increase abortive product formation, indicating there is a correlation between scrunching and the position of the wild-type TSS and that scrunching may facilitate promoter escape.

Results

The TSS of *rrnB* P1 Is Not Dependent on Relative Nucleotide Concentrations in Vitro. Although *rrnB* P1 is a very strong promoter, it differs from consensus in the –35 hexamer, has suboptimal spacing between the –10 and –35 hexamers, lacks a consensus extended –10 element or discriminator sequence (Fig. 1 *A* and *B*), and forms very unstable open complexes (15, 16). These promoter features result in a requirement for higher initiating NTP concentrations for transcription initiation than at most promoters, allowing regulation of rRNA production by changing NTP levels when they are not saturating (17–19). To investigate whether the NTP concentration contributes to the unusual TSS at *rrnB* P1, we carried out in vitro transcription at a range of NTP concentrations and mapped the TSS by primer extension (Fig. 1*C*). When a high concentration of all four NTPs (500 μM A, G, C, U) was used, wild-type *rrnB* P1 initiated predominantly with ATP 9 bp from the downstream end of the –10 hexamer (referred to as +9A), with minor percentages initiating at +8C (17%), +7C (12%), or +6A (8%) (Fig. 1 *A* and *C*, lane 9). A similar distribution of TSSs was observed when any single NTP was used at 10-fold higher concentration than the others (500 μM versus 50 μM; Fig. 1*C*, lanes 5–8). These data suggest that although the efficiency of initiation at *rrnB* P1 is unusually dependent on the NTP concentration, TSS position at this promoter is relatively insensitive to the ratio of NTP concentrations available.

Mutations Throughout *rrnB* P1 Alter the TSS. To determine how promoter sequence influences TSS selection, we examined a preexisting library of 38 *rrnB* P1 promoter variants for those that displayed an altered TSS in vitro. This library included some site-directed mutations in *rrnB* P1, as well as mutations isolated in an in vitro selection for promoters that formed more stable *rrnB* P1–RNAP complexes and in screens for loss of regulation (16, 19).

Of the 38 promoter mutants tested, 18 displayed an altered TSS, in most cases shifting the majority of starts from +9A to +6A (Fig. 2 *B* and *C*). (TSSs at +7 or +8 were much less frequent, because pyrimidines are disfavored for initiation, whereas +9A and +6A TSSs are preceded by a preferred pyrimidine). Several *rrnB* P1 variants exhibited pronounced shifts from +9A to +6A, including three previously characterized mutations C-7G, C-16G/C-17T, and insC-18. Of these variants, the one with the most complete shift was the C-7G promoter (Fig. 2*B*, lane 6). (Promoter positions are numbered with respect to the wild-type TSS, referred to as +1; Fig. 2*A* and *B*.) The C-7G substitution creates a consensus G at this discriminator position, improving promoter interactions with $\sigma_{1.2}$ (16); the C-16G/C-17T mutation creates a consensus extended –10 element (promoter referred to as Ext –10), improving interactions with $\sigma_{3.0}$ (20); and the insC-18 mutation increases the spacing between the –10 and –35 hexamers to the consensus 17 bp (Fig. 1 *A* and *B*).

All three of these promoter mutations greatly increased the lifetime of open complexes (16). In contrast, two promoter variants also made very stable open complexes but did not shift the TSS, *rrnB* Dis = C-5A/G-6T/C-7A, and T-33A (Fig. 2*B*, lane 7, and Fig. 2*C*). *rrnB* Dis creates a more A+T-rich discriminator, and T-33A creates a consensus –35 hexamer. These promoters initiated at +9A, like the wild type (Fig. 2*B*, lane 7, and Fig. 2*C*), indicating that open complex stability per se is insufficient to explain the change in TSS (16).

Other *rrnB* P1 mutations also shifted the TSS from +9A to +6A, including several that increased the length between the –10 and –35 hexamers to 17, 18, or 19 bp or that weakened the –35 element interaction with σ_4 (e.g., G-34A, T-35C, or T-36C). Five TSS-altering mutations mapped to the start site region of the promoter, creating tranversions at 7, 8, or 9 bp downstream of the –10 element. Some of these changes in TSS correlated with creation or disruption of a favored pyrimidine–purine sequence at –1/+1. For example, the C-1A mutation created a preferred purine preceded by a pyrimidine at a more preferred start position (5′-CA-3′, 7 and 8 bp downstream from the –10 hexamer) and disrupted the 5′-CA-3′ sequence 8 and 9 bp downstream from the –10 element, switching the TSS from +9 to +8.

The Template Strand Is Positioned Differently in *rrnB* P1 Promoter Complexes That Initiate at +6 Rather than +9. To investigate the mechanism(s) responsible for TSS selection, we monitored the position of the template strand with respect to the RNAP active site using representative mutant or wild-type promoters with different TSSs: *rrnB* C-7G (+6A), wild-type *rrnB* P1 (+9A), and *rrnB* Dis (+9A). Open complexes were formed with radiolabeled promoter fragments and RNAP in which the active site Mg^{2+} was

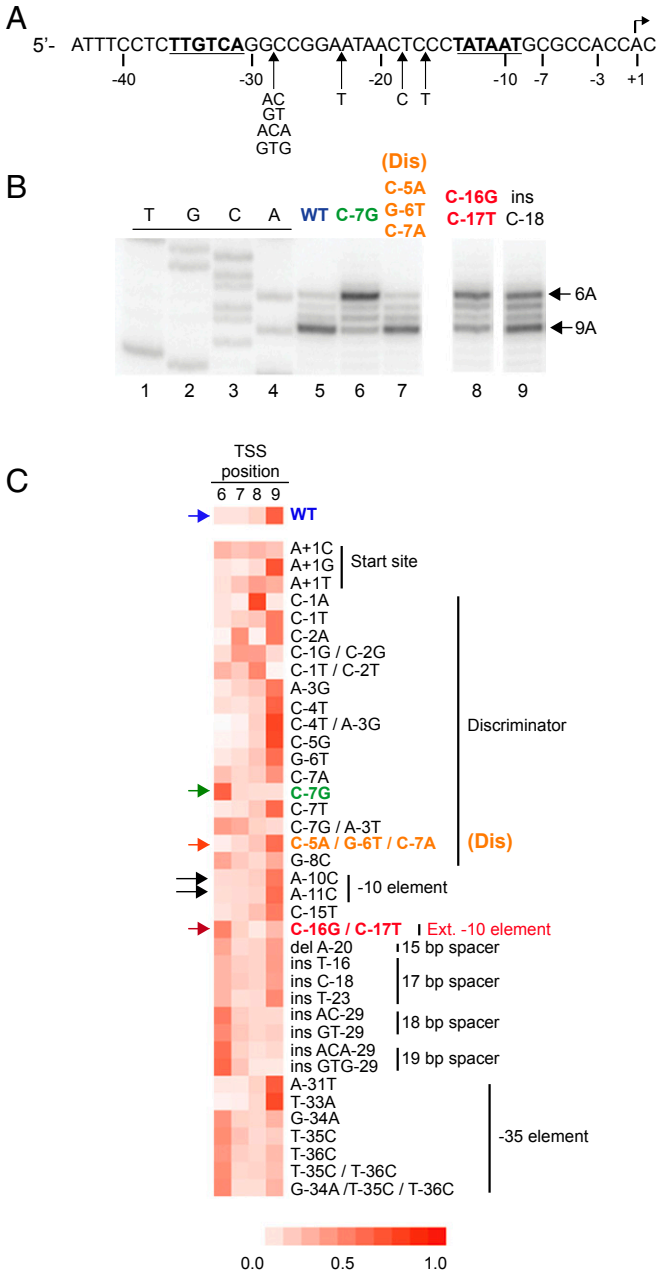

Fig. 2. Some *rrnB* P1 promoter variants shift the TSS. (*A*) *rrnB* P1 promoter sequence. Positions of insertion mutants shown in *C* are indicated below the sequence. The –35 and –10 elements are in bold, and the TSS is indicated with an arrow. (*B*) TSSs from wild-type or the indicated mutant *rrnB* P1 promoters determined by primer extension of in vitro transcripts. A representative gel is shown. TSS positions at +9A or +6A are indicated. Lanes 8 and 9 are from the same gel as lanes 1–7, but some intervening lanes have been omitted for clarity. (*C*) Heat map depicting the TSS distribution of *rrnB* P1 variants. Promoter names are indicated. Columns represent unique TSS positions 6, 7, 8, or 9 bp downstream of the –10 hexamer, and color intensity represents the fraction of total transcription initiating at each position (key is at the bottom). Values are the average of three independent experiments. Differences in overall promoter activity are not depicted. Color-coded arrows indicate variants discussed in other figures. Black arrows indicate mutations (A-10C, A-11C) that were used to test for potential utilization of an alternative –10 hexamer.

replaced with Fe^{2+}. Hydroxyl radicals, generated by the Fe^{2+} RNAP, cleaved template strand DNA positions in close proximity to the active site. The majority of strand cleavage in complexes with the wild-type and *rrnB* Dis promoters occurred 7 or 8 bp downstream from the –10 hexamer. However, with *rrnB* P1 C-7G, there was a pronounced shift in cleavage distribution to include positions 5–8 bp downstream from the –10 element (Fig. 3 *A* and *B*). We conclude that the C-7G mutation affects the position of the template strand with respect to the RNAP active site and that strand placement occurs before, and is independent of, the addition of NTPs during initiation.

Interactions Between σ^{70} Region 1.2 and the Discriminator Nontemplate Strand Contribute to Positioning the Template Strand. To identify whether the nontemplate strand G or the template strand C (or both) was responsible for repositioning the template strand and shifting the TSS of the C-7G promoter, start sites were determined for promoter templates with a single bp mismatch at –7 (Fig. 3*C*). The start site distribution for a bubble template with a nontemplate strand G and a template strand A (G/A bubble) was similar to that for the C-7G promoter, whereas the distribution for the A/C bubble template was similar to that of wild-type *rrnB* P1 (Fig. 3 *D* and *E*). These data support the model that the nontemplate strand G rather than the template strand C accounts for the TSS shift with the C-7G promoter and indicate that the identity of the –7 position on the nontemplate strand, not the G+C content of the discriminator, is responsible for these effects.

We showed previously that the nontemplate strand G of the C-7G promoter interacts with $\sigma_{1.2}$ (16), and subsequent genetic studies (21) and structural studies (22) showed that σ residues M102 and R103 contribute to this interaction. We therefore compared the TSSs produced by the wild-type RNAP and σ M102A/R103A mutant RNAP and by the C-7G and *rrnB* P1 Dis promoters (Fig. 3 *F* and *G* and Fig. S1). Consistent with the hypothesis that the $\sigma_{1.2}$ interaction with the nontemplate strand G at –7 is important for TSS selection, the C-7G mutation largely prevented shifting of the TSS when the promoter was transcribed with the σ M102A/R103A RNAP. Individual substitutions in several neighboring residues in $\sigma_{1.2}$ also reduced shifting of the TSS on the C-7G promoter (Fig. S1). In contrast, the start site distribution for the *rrnB* Dis promoter was not affected by the $\sigma_{1.2}$ M102A/R103A substitutions, consistent with the lack of a consensus G residue at nontemplate strand position –7 in this promoter.

Models for Template Strand Positioning. We considered several possible mechanisms to account for the differences in template strand position observed with the wild-type and mutant *rrnB* P1 promoters, including recognition of an alternative –10 hexamer, "sliding" of promoter DNA downstream of the –35 hexamer relative to RNAP, and "scrunching" of additional downstream DNA into RNAP in the open complex.

An alternative –10 hexamer sequence TAATGC (–12 to –7 in wild-type *rrnB* P1 numbering) is spaced 18 bp downstream from the –35 hexamer. Recognition of this element by RNAP rather than the TATAAT (–14 to –9) would allow transcription to initiate at the observed +9A position (wild-type promoter numbering) at a distance of 7 bp from the end of the –10 hexamer. If this alternative –10 hexamer were used by RNAP, then substitution for the A at –11 (Fig. 2*C,* black arrow), the critical second position in the –10 hexamer, would be expected to eliminate stable recognition. However, this was not observed: An A-11C (or an A-10C) substitution did not prevent initiation at +9A and had no effect on TSS distribution, strongly suggesting the alternative –10 hexamer explanation for the +9A TSS is incorrect.

In the sliding model, strain in the wild-type *rrnB* P1 open complex, imposed by the short (16 bp) spacing between the –10 and –35 recognition elements, destabilizes σ interactions with the –10 element region, resulting in an open complex anchored only

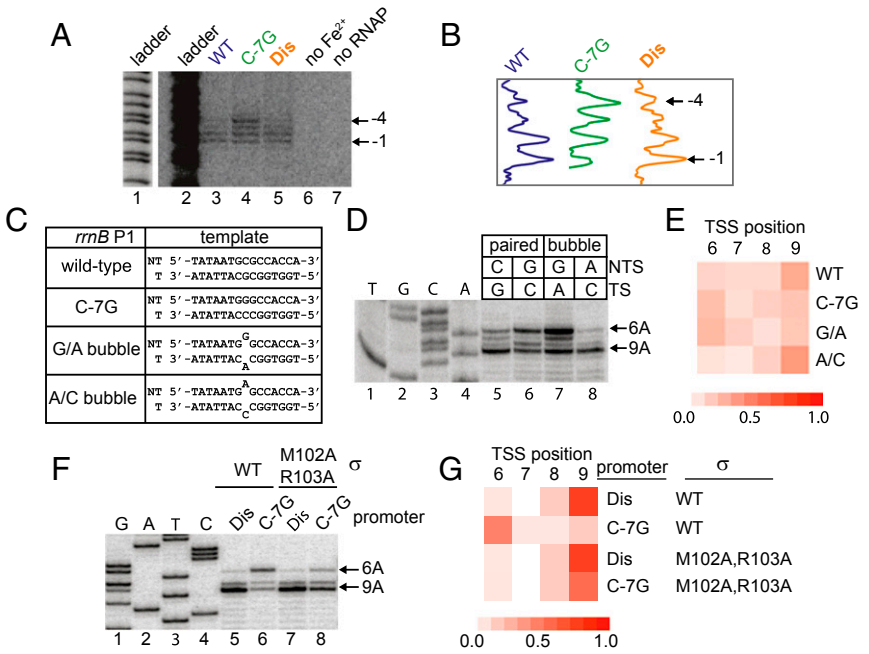

Fig. 3. Interactions between the nontemplate strand of the discriminator and $\sigma_{1.2}$ alter template strand position in the open complex. (*A*) Active site-specific Fe^{2+}-mediated DNA cleavage in open complexes at the wild-type, C-7G, or Dis promoters (*Materials and Methods*). A+G sequence ladder labeled lane 1 is the same as gel lane 2 but at a lower signal intensity. (*B*) Traces of lanes 3–5 in gel shown in *A*. (*C*) Nontemplate (NT) and template (T) strand DNA sequences from −14 to +1 in *rrnB* P1 promoters used in panels *D* and *E*. Bubble templates contain a G/A or an A/C mismatch at position −7. (*D*) TSSs from fully paired wild-type or C-7G promoter templates or from bubble templates with a −7 mismatch, determined by primer extension from transcripts generated in vitro. Identity of the −7 position on each strand is indicated. (*E*) Heat map (as in Fig. 2*C*) representing the TSS distribution of fully paired or bubble templates in *D*. Values are the average of three independent experiments. (*F*) Start sites of transcription from the *rrnB* Dis or C-7G promoters with wild-type RNAP or σM102A/R103A RNAP determined by primer extension of transcripts generated in vitro. (*G*) Heat map representing TSS distributions in *F*.

by $\sigma_{4.2}$ interactions with the −35 element. Promoter DNA downstream of the −35 element then "slides" with respect to RNAP, repositioning the 13 bp open complex bubble downstream (i.e., shifting the register of the DNA with respect to the active site of the enzyme), explaining why transcription initiates at position +9A rather than +6A (wild-type *rrnB* P1 numbering).

Effects of all of the *rrnB* P1 promoter mutations on TSS selection are consistent with the sliding model. For example, spacer insertions that would relieve the strain imposed by the 16 bp spacer and increase −10 region interactions, preventing sliding, shifted the TSS to +6. Mutations in the −35 hexamer that would disrupt interactions with sigma 4.2, thereby relieving the strain imposed by the 16 bp spacing, also shifted the TSS to +6. However, the loss of critical −10 hexamer interactions with the nontemplate strand (23) in the hypothetical shifted complex predicted by the sliding model would be expected to result in DNA bubble collapse rather than bubble reformation downstream, making this model an unlikely explanation for the +9A TSS. Furthermore, this model is inconsistent with the results of cross-linking experiments (see rrnB *P1 Open Complexes Are Scrunched*).

The third model, open complex scrunching, proposes that the extra DNA length in the *rrnB* P1 transcription bubble between the TSS and the −10 element is accommodated within the RNAP main channel, as has been observed in initial transcribing complexes (24). In contrast to the situation in initial transcribing complexes, however, scrunching of the open complex would occur before NTP addition, driven by binding free energy instead of by NTP hydrolysis, and effects of promoter mutations that changed the TSS from +9A to +6A would derive from strengthening of contacts with σ and prevent open complex scrunching. We suggest below that weakening, but not eliminating, interactions with the −10 region results in scrunching and the unusual *rrnB* P1 TSS.

***rrnB* P1 Open Complexes Are Scrunched.** We tested these models by using a cross-linking approach (24) to map the path of DNA at amino acid and nucleotide resolution in open complexes formed on promoters with different TSSs (*rrnB* P1, *rrnB* Dis, *rrnB* Ext −10, and *rrnB* C-7G). The photoactivatable amino acid Bpa (25) was incorporated into RNAP one residue at a time at amino acid locations chosen because they were shown previously to cross-link to DNA in promoter complexes (24). After formation of *rrnB* P1 promoter complexes and UV irradiation, the resulting open complex cross-links to DNA were mapped to nucleotide resolution using primer extension.

Each of the models predicted that Bpa incorporated at the leading edge of RNAP would make cross-links to DNA further downstream on the promoters that initiated at +9A than on the promoters that initiated at +6A. Consistent with this prediction, and with the hydroxyl radical cleavage patterns shown in Fig. 3*A*, the major cross-links from residues β′R1148-Bpa and β′K1170-Bpa (which interact with the downstream duplex) were 2–3 bp further downstream on the wild-type and *rrnB* Dis promoters than on the *rrnB* Ext−10 and *rrnB* C-7G promoters (Fig. 4*A*, panels 3 and 4; summarized in Fig. 4*B*).

In contrast, the models predicted different results for cross-links from RNAP to positions in the spacer and −10 hexamer regions of the promoter. The sliding and alternate −10 hexamer models predicted that these cross-links would be at different positions for promoters with different TSSs, whereas the open complex scrunching model predicted that the positions of these cross-links would not change for promoters with different TSSs. As shown in Fig. 4*A*, panels 1 and 2, in both sets of promoters the major cross-link from β′T48-Bpa was to −23/−24 in the template strand spacer, and the major cross-link from σR465-Bpa was to position −12, at the upstream edge of the −10 hexamer. These results are consistent only with the open complex scrunching

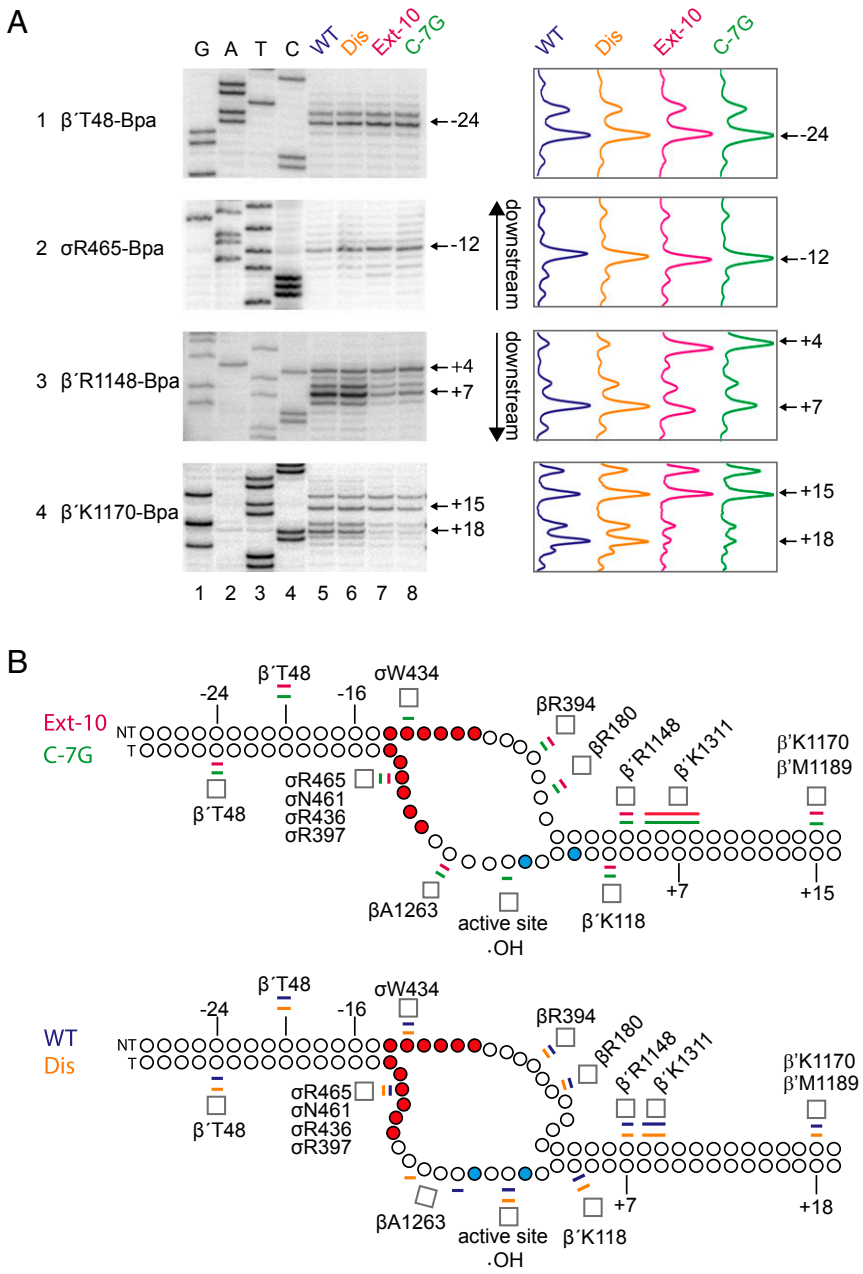

Fig. 4. *rrnB* P1 forms scrunched open complexes. (*A*) Cross-links from β′T48-Bpa or σR465-Bpa to the template strand or from β′R1148-Bpa or β′K1170-Bpa to the nontemplate strand in open complexes formed with the wild-type *rrnB* P1 (lanes 5), Dis (lanes 6), Extended −10 (lanes 7), or C-7G (lanes 8) promoters. Cross-link positions were determined by comparison of primer extension products from cross-linked complexes to GATC sequence ladders generated from *rrnB* Dis (lanes 1–4) using the same strand-specific primers. Traces of lanes 5–8 are shown at the right of the gel images. Bpa-containing RNAPs were tested to ensure Bpa incorporation did not affect TSS (Fig. S4). (*B*) Schematic of DNA strands in open complexes at the extended −10 and C-7G promoters (*Top*; nonscrunched) or the wild-type and Dis promoters (*Bottom*; scrunched). Positions of cross-links or sites of template strand cleavage by Fe^{2+} in the RNAP active site (•OH) from data in Figs. 3*A* and 4*A* and Figs. S2 and S3 are indicated. The −10 hexamer positions are in red, and template strand thymines at +6 and +9 are in cyan. Promoter numbering is for wild-type *rrnB* P1. Colored bars above and below each strand represent the most highly cross-linked position or range of positions for each promoter (wild-type *rrnB* P1, blue; Dis, orange; Ext −10, pink; C-7G, green). Cross-link positions in the downstream duplex are shifted farther downstream, and the downstream edge of the bubble region is expanded in the scrunched open complexes relative to the nonscrunched complexes (*Bottom* panel vs. *Top* panel). Note that β′T48-Bpa cross-links to both the template strand (Fig. 4*A*) and the nontemplate strand (Fig. S2).

model, in which an increase in the number of nucleotides in the bubble results from pulling extra DNA into the enzyme in promoter complexes initiating at +9, and DNA positions upstream of the nucleotides in the expanded part of the bubble do not shift relative to RNAP.

To determine the locations of specific DNA positions in the four promoter complexes more precisely, we mapped cross-links from seven additional Bpa-containing RNAPs that cross-linked to the nontemplate strand (σT552, β′T48, σW434, βR394, βR180, β′K1311, and β′M1189) and from five additional Bpa-RNAPs that cross-linked to the template strand (σR436, σR397, σN461, βA1263, and β′K118) (Figs. S2 and S3). Cross-links from RNAP to positions upstream of −11 on the template strand and −4 on the nontemplate strand generally mapped to the

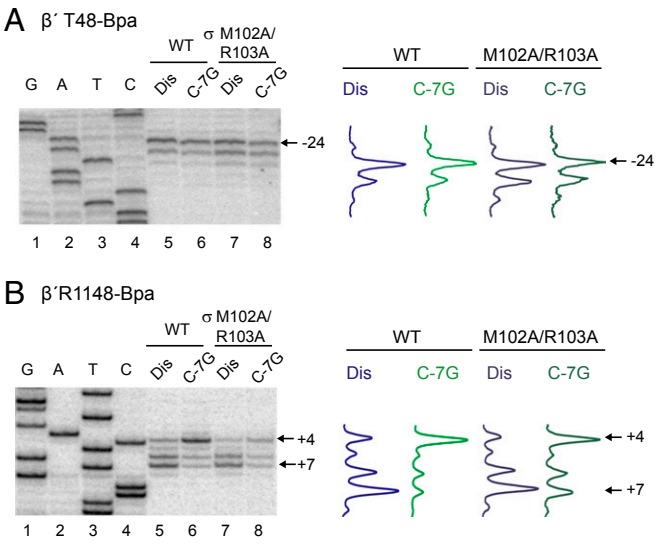

Fig. 5. Discriminator interactions with $\sigma_{1.2}$ inhibit scrunching. Cross-links from β′T48-Bpa to the template strand (*A*) or from β′R1148-Bpa to the nontemplate strand (*B*) in open complexes formed with wild-type RNAP or σM102A/R103A RNAP and either the Dis or C-7G promoters. Cross-link positions were determined by primer extension from strand-specific primers (lanes 5–8) and are compared with GATC sequencing ladders (lanes 1–4) generated from the Dis promoter with the same primers. Traces for lanes 5–8 are shown next to the gel.

same positions on the four promoters, whereas cross-links from residues in RNAP to DNA downstream of the bubble were generally shifted downstream on the wild-type promoter and *rrnB* Dis compared with cross-links to *rrnB* Ext –10 and *rrnB* C-7G. The magnitude of these downstream shifts is similar to those observed previously in an initial transcribing complex that contained a 5-mer RNA (24).

$\sigma_{1.2}$ Interactions with –7G on the Nontemplate Strand Prevent Open Complex Scrunching. The C-7G promoter mutation, which improves interactions with $\sigma_{1.2}$ (16, 21), exhibited the most pronounced shift in TSS (to +6A; Fig. 2), and mutations in $\sigma_{1.2}$ reduced the efficiency of this shift (σM102A/R103A; Fig. 3 *F* and *G*). These results predicted that the σM102A/R103A RNAP would show open complex scrunching even on the C-7G promoter by weakening the –7G interaction with $\sigma_{1.2}$.

Mapping of cross-links in the open complex confirmed the prediction. The template strand β′T48-Bpa cross-link to *rrnB* Dis and C-7G was similar for RNAPs with wild-type σ or $\sigma_{1.2}$ M102A/R103A (to –23, –24; Fig. 5*A*), indicating that the interface between RNAP and the upstream duplex DNA was not altered by the σM102A/R103A substitutions. In contrast, the downstream cross-link between the β′R1148-Bpa and the C-7G promoter shifted from +4 with wild-type RNAP to include +6 and +7 with the σM102A/R103A RNAP. Thus, despite the presence of the consensus nontemplate strand G in the C-7G promoter, the σM102A/R103A substitutions weakened the discriminator–$\sigma_{1.2}$ interaction, increasing the fraction of scrunched complexes in which +9A on the template strand is adjacent to the active site in RNAP.

Abortive RNA Production Is Reduced from Promoters That Form Scrunched Open Complexes. We proposed previously that the template strand bulge caused by scrunching during initial transcription increases the rate of promoter escape (24). We addressed the potential relationship between open complex scrunching and productive transcription by measuring the level of abortive RNA products synthesized and the productive RNA-to-abortive ratio by the wild-type, *rrnB* Dis, *rrnB* Ext –10, and *rrnB* C-7G promoters in the presence of all four NTPs (Fig. 6*A*). High levels of abortive RNAs (RNAs shorter than ~12 nt) generally reflect inefficient promoter escape (11). The abortive RNA level was much lower, and the productive-to-abortive RNA ratio was much higher for the promoters that made scrunched open complexes, wild-type *rrnB* P1 and *rrnB* Dis, compared with the promoters that did not make scrunched open complexes, *rrnB* Ext –10 and *rrnB* C-7G (Fig. 6*A*), consistent with the model that initiating transcription from a scrunched state increases promoter efficiency (Fig. 6*B*).

Discussion

We took advantage of the unusual TSS of *rrnB* P1 to study the determinants of TSS selection. First we identified a set of *rrnB* P1 mutations that led to a shift in TSS from +9A to +6A. Then

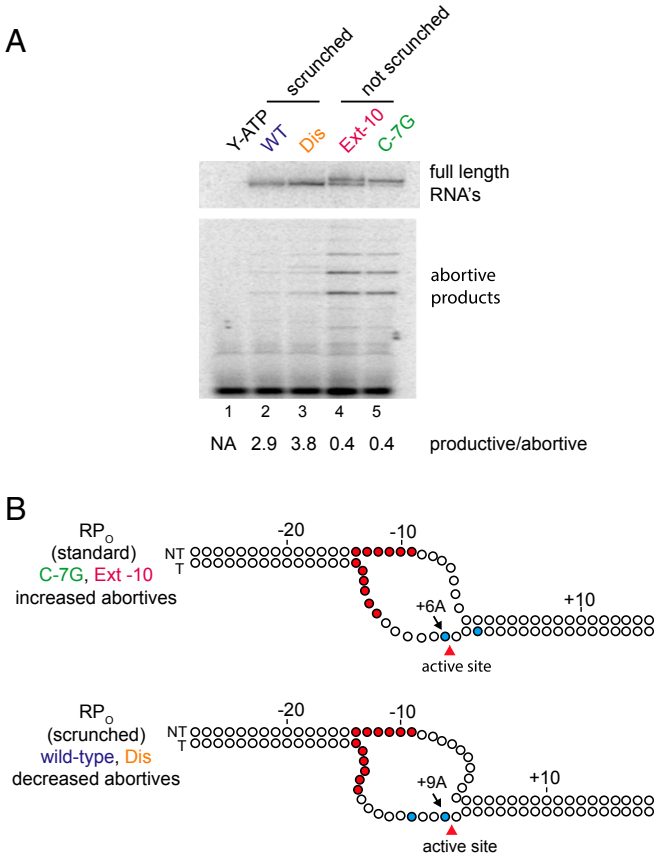

Fig. 6. Role of open complex scrunching in promoter escape. (*A*) RNAs generated by transcription in the presence of γ[^{32}P]ATP from the wild-type *rrnB* P1, Dis, Ext –10, or C-7G promoters were electrophoresed on a 20% acrylamide gel to resolve abortive transcripts. Panels show full-length (productive) and abortive transcripts (estimated as 4–8 nt in length) from the same gel, but the region of the image between the long and short RNAs has been eliminated to save space. The ratio of productive (full-length) to abortive transcripts is indicated below each lane. Relative levels of productive transcripts for each promoter are as follows: lane 2 wild-type, 1.0; lane 3 Dis, 1.8; lane 4 Ext –10, 1.6; lane 5 C-7G, 1.1. (*B*) Open complex scrunching increases promoter efficiency. Positions in the –10 hexamer are indicated by filled red circles. The positions of the two alternative start sites (at +6A or +9A from the –10 hexamer) are shown as filled cyan circles. The active site in RNAP adjacent to the TSS is indicated by a red arrowhead. (*Top*) C-7G or Ext –10 promoter, with the TSS at +6 (arrow); the nonscrunched configuration leads to abortive transcripts. (*Bottom*) Wild-type or Dis promoters with the TSS at +9 (arrow); the scrunched configuration leads to fewer abortive transcripts.

we used a cross-linking approach to compare the paths of DNA with respect to RNAP in open complexes that initiate predominantly at +9A versus +6A and found the TSS position was diagnostic of the presence/absence of open complex scrunching. Finally, we examined the effects of representative mutations from the two promoter classes on abortive product formation as an indicator of their effects on promoter escape.

Promoter Properties Leading to a Scrunched Open Complex. The wild-type *rrnB* P1 promoter makes unstable open complexes, likely because the shorter than consensus spacing between the –10 and -35 hexamers creates strain that reduces interactions between σ and promoter DNA in and around the –10 element. *rrnB* P1 promoter mutations that led to stronger interactions between $\sigma_{1.2}$ or $\sigma_{3.0}$ (with the discriminator region or extended –10 element, respectively) resulted in a shift from a +9 TSS (WT) to +6 (mutants), a signature of the loss of scrunching. We therefore suggest that open complex scrunching requires weak sigma interactions; strong contacts with the –10 region constrain scrunched open complex formation.

Not only mutations in the vicinity of the –10 region itself (e.g., in the discriminator or the extended –10 element) but also well upstream of the –10 region shifted the TSS to +6 and reduced scrunching. We suggest that mutations that increase –10/–35 spacer length (to 17, 18, or 19 bp), or mutations away from consensus in the –35 element, shift the TSS and reduce scrunching by reducing the strain imposed by the 16 bp spacing between the –10 and –35 elements, thereby stabilizing σ interactions with the promoter in the vicinity of the –10 element.

We note that the *rrnB* Dis promoter also forms a more stable complex, but it creates an A+T-rich sequence downstream of the –10 that likely facilitates strand opening rather than strengthens interactions with $\sigma_{1.2}$. This is consistent with the observation that it does not interfere with scrunching.

We suggest that weakening –10 region contacts (or a subset of contacts) alters the local environment in the main channel, allowing for accommodation of a longer than usual transcription bubble within the channel. Our cross-linking data support the presence of an expanded bubble within the channel, with extra nt accommodated between the downstream duplex and position –4 on the nontemplate strand and between the downstream duplex and –11 on the template strand (Fig. 4). Cross-links to positions upstream of the scrunched region were the same for promoters with TSSs at +6 and +9, distinguishing the scrunched open complex model from the sliding model described in *Results*.

In separate studies in which all possible DNA sequences between the –10 hexamer and the TSS were examined in vitro, as well as the distance between the leading edge and trailing edge of the promoter complex, it was concluded that TSS position correlates with scrunching state and is dependent on the absence of a $\sigma_{1.2}$ interaction with the discriminator (26, 27). These data are consistent with the results reported here.

Open Complex Intermediates and Open Complex Scrunching. There are multiple kinetically significant intermediates on the path to open complex formation, and the relative occupancy of those intermediates depends on the overall promoter sequence. At least two different promoter open complexes have been characterized in studies of the λP_R promoter, an initial unstable complex, called I_2, and the more familiar stable open complex, RP_O (28, 29). In the unstable λP_R open complex, the DNA strands are melted from –11 to +2, and the template strand is positioned in the active site, but the nontemplate strand is not yet in its final position. In the stable λP_R open complex, additional RNAP–promoter contacts are made: The nontemplate strand is positioned in the cleft, and downstream DNA contacts with the β′ jaw and clamp region are established (28, 29).

rRNA promoters do not form a long-lived complex corresponding to the stable open complex formed by most other promoters (16, 30). Record and colleagues (28, 29) noted that there may be some similarities between the scrunched *rrnB* P1 open complex and the I_2 intermediate formed by λP_R. Our findings suggest that because of its less stable contacts with σ, the *rrnB* P1 scrunched complex, like λP_R I_2, might be more flexible than the more stable RP_O characterized at other promoters. Binding of the initial NTP to the scrunched *rrnB* P1 open complex not only would lead to a TSS at +9 but also would result in an RNA synthesizing conformation without prior population of the more stable conformation that characterizes standard open complexes.

Open Complex Scrunching and Abortive Product Formation. The previously described scrunched complex, formed during initial RNA synthesis, is a high-energy intermediate requiring NTP addition that is thought to contribute to the energy needed to break interactions between RNAP and the promoter, facilitating promoter escape and productive RNA synthesis (9, 10). In complexes containing ∼5 nt of RNA, we proposed that scrunched template strand DNA could play a direct role in promoter escape by helping displace $\sigma_{3.2}$ from the RNA exit channel, either by impinging on $\sigma_{3.2}$ directly or by destabilizing the complex indirectly by disruption of interactions between $\sigma_{3.2,}$ the β′ clamp, and the β flap (24).

In contrast, our data indicate that the scrunched open complex at *rrnB* P1 is not a high-energy intermediate. Although it is not kinetically stable, it is energetically favored relative to the nonscrunched complex. There might be a lower energetic cost associated with scrunching an open complex than an initial transcribed complex, because there is no need to accommodate a growing RNA chain, making more space available in the main channel for the expanded bubble. The lack of a large energetic barrier would reduce abortive synthesis, thereby facilitating the transition to the elongation phase of transcription. In this way, the scrunched open complex might already be on the path to displacement of $\sigma_{3.2}$, "primed" to release RNAP from the promoter, bypassing the requirement for creating strain during early RNA synthesis.

A scrunched open complex that initiated at +9 (2 bp further downstream from the –10 hexamer than a standard promoter complex) would have more scrunched DNA in its main channel during initial transcription than a standard complex. Therefore, during initial RNA synthesis, for a given RNA length, the extra scrunched DNA in rRNA promoters might generate more force on the RNA exit channel, better inhibit abortive product formation, and facilitate escape more efficiently than standard promoters.

Implications for rRNA Promoter Evolution. At high growth rates, rRNA synthesis accounts for the majority of RNAP activity in the cell to keep up with the high demand for ribosome synthesis (31). The extraordinary strength of *rrn* P1 promoters derives from an extremely high association rate with RNAP (32–34) but also from making few or no abortive transcripts—that is, a high productive-to-abortive RNA ratio. Our data are thus consistent with a model in which open complex scrunching contributes to rRNA promoter efficiency (Fig. 6).

Promoter mutations that affect *rrnB* P1 interactions with σ strongly affect open complex formation, complicating assumptions about promoter escape from measurements of promoter activity alone. For example, kinetic measurements of the *rrnB* P1 C-7G promoter indicated that it associates ∼11-fold faster with RNAP than the wild-type *rrnB* P1 promoter in vitro (16). However, the C-7G and wild-type promoters make approximately the same amount of full-length RNA in vitro (Fig. 6*A*), and the C-7G promoter is less than twofold more active than the wild-type promoter in vivo at high growth rates (i.e., when they are not

being inhibited by regulatory factors) (16). These results are consistent with the model that scrunching during open complex formation and initial transcription helps compensate for the weak interactions with σ in the wild-type promoter complex.

Regulation of rRNA synthesis by small molecule transcription factors (NTP concentration, ppGpp, DksA) is dependent on weak interactions of the promoter with σ regions 1.2 and 3.0 (15, 16). As we have proposed previously (15, 35), to compensate for weak core promoter interactions, rRNA promoters have evolved alternative mechanisms to help recruit RNAP, namely UP element interactions with the α subunit C-terminal domain (33) and activation by the transcription factor Fis (36). Here we suggest that by reducing abortive product formation, open complex scrunching also contributes to high rRNA synthesis without compromising promoter regulation. Our results thus fully agree with the model that promoter strength and regulation are the result of an optimization process, with tradeoffs involving multiple kinetic parameters (37).

Materials and Methods

Details for all procedures are provided in *SI Materials and Methods*.

Plasmids and Oligonucleotides. Plasmids are listed in Table S1 and oligonucleotides in Table S2.

Purification of RNAPs. Native *E. coli* RNAP holoenzyme and core enzyme were purified as described (38). Wild-type or mutant His-tagged σ^{70} subunits were purified from BL21DE3 containing pET28-derived overexpression plasmids by Ni-agarose affinity chromatography. Bpa-containing core RNAPs were purified from BL21DE3 containing multisubunit overexpression plasmids encoding TAG codons at specific sites and a Bpa-specific suppressor tRNA/tRNA synthetase plasmid (25) as described (24). Bpa-containing σ^{70} subunits were similarly purified from DH10B containing pBAD24 derivatives encoding σ^{70} with TAG codons at specific sites.

In Vitro Transcription and Primer Extension Mapping of TSSs. Multiple-round in vitro transcription of the *rrnB* P1 promoter or promoter variants was carried out with supercoiled plasmid DNA templates and wild-type or σ^{70} M102A/R103A *E. coli* RNAP. TSSs were determined by primer extension with Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase using a ^{32}P 5′-end–labeled primer annealing ∼100 nt downstream of the RNA 5′-end, and products were separated on 11% (wt/vol) acrylamide–urea gels. Sequence markers were generated by ThermoSequenase Cycle Sequencing using the same primer. Linear bubble templates (171 bp) for in vitro transcription (Fig. 3) contained a single base mismatch at promoter position –7 and were created by denaturation and annealing of biotinylated and nonbiotinylated PCR products generated from wild-type and mutant promoters and separation of the fully nonbiotinylated from biotinylated products (16). Templates for analysis of short abortive products (Fig. 6) were linear wild-type or mutant *rrnB* P1 promoter fragments, and transcripts generated in the presence of γ[^{32}P]ATP were analyzed on 20% acrylamide–urea gels.

Hydroxyl Radical Cleavage of Template Strand DNA in Open Complexes. Open complexes were formed with linear *rrnB* P1 wild-type or mutant promoter fragments ^{32}P-labeled at the 5′-end of the template strand. Strand cleavage was initiated by replacement of Mg^{2+} in the RNAP active site by Fe^{2+} and generation of hydroxyl radicals at the bound Fe^{2+}, as described (39). Cleavage products were separated on 11% acrylamide–urea gels.

Cross-Linking and Primer Extension Mapping of Cross-Links. Open complexes were formed with Bpa-containing RNAPs and wild-type or mutant *rrnB* P1 promoter-containing plasmids at 37 °C and then irradiated with 365 nM UV light. Promoter positions cross-linked to RNAP were determined by primer extension with template or nontemplate strand-specific end-labeled primers, as described (24). Primers anneal to the strand complementary to the cross-linked strand, and extension of the primer terminates at the cross-linked position. Cross-linked positions were identified by comparison with sequence ladders generated with the same primers after electrophoresis on 9.5% acrylamide–urea gels.

ACKNOWLEDGMENTS. We thank Peter Schultz for the pSupT/BpF plasmid; members of the R.L.G. laboratory for critical reading of the manuscript; members of the R.L.G., Landick, Wassarman, and Record laboratories for insightful discussions; Chris Lennon and Tamas Gaal for purified RNAP; Shani Haugen and Tamas Gaal for *rrnB* P1 promoter and σ region 1.2 variants; and Ethan Grefe for assistance in making *rrnB* P1 promoter mutations. J.T.W. was supported in part by a Molecular Biosciences Training Grant from the NIH and a fellowship from the Department of Bacteriology. P.C. was supported in part by an NIH Genetics Training Grant. Work in the R.L.G. laboratory is supported by NIH Grant R37 GM37048.

1. Liu J, Turnbough CL, Jr (1994) Effects of transcriptional start site sequence and position on nucleotide-sensitive selection of alternative start sites at the *pyrC* promoter in *Escherichia coli*. *J Bacteriol* 176(10):2938–2945.
2. Lewis DE, Adhya S (2004) Axiom of determining transcription start points by RNA polymerase in *Escherichia coli*. *Mol Microbiol* 54(3):692–701.
3. Han X, Turnbough CL, Jr (2014) Transcription start site sequence and spacing between the -10 region and the start site affect reiterative transcription-mediated regulation of gene expression in *Escherichia coli*. *J Bacteriol* 196(16):2912–2920.
4. Basu RS, et al. (2014) Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J Biol Chem* 289(35):24549–24559.
5. Harley CB, Reynolds RP (1987) Analysis of *E. coli* promoter sequences. *Nucleic Acids Res* 15(5):2343–2361.
6. Thomason MK, et al. (2015) Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J Bacteriol* 197(1):18–28.
7. Turnbough CL, Jr, Switzer RL (2008) Regulation of pyrimidine biosynthetic gene expression in bacteria: Repression without repressors. *Microbiol Mol Biol Rev* 72(2):266–300.
8. Robb NC, et al. (2013) The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: Implications for transcription start-site selection. *J Mol Biol* 425(5):875–885.
9. Kapanidis AN, et al. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314(5802):1144–1147.
10. Revyakin A, Liu C, Ebright RH, Strick TR (2006) Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* 314(5802):1139–1143.
11. Vo NV, Hsu LM, Kane CM, Chamberlin MJ (2001) In vitro studies of transcript initiation by *Escherichia coli* RNA polymerase. 3. Influences of individual DNA elements within the promoter recognition region on abortive initiation and promoter escape. *Biochemistry* 42(13):3798–3811.
12. Ginsburg D, Steitz JA (1975) The 30 S ribosomal precursor RNA from *Escherichia coli*. A primary transcript containing 23S, 16S, and 5S sequences. *J Biol Chem* 250(14):5647–5654.
13. Borukhov S, Sagitov V, Josaitis CA, Gourse RL, Goldfarb A (1993) Two modes of transcription initiation in vitro at the *rrnB* P1 promoter of *Escherichia coli*. *J Biol Chem* 268(31):23477–23482.
14. Josaitis CA, Gaal T, Gourse RL (1995) Stringent control and growth-rate-dependent control have nonidentical promoter sequence requirements. *Proc Natl Acad Sci USA* 92(4):1117–1121.
15. Haugen SP, Ross W, Gourse RL (2008) Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat Rev Microbiol* 6(7):507–519.
16. Haugen SP, et al. (2006) rRNA promoter regulation by nonoptimal binding of sigma region 1.2: An additional recognition element for RNA polymerase. *Cell* 125(6):1069–1082.
17. Gaal T, Bartlett MS, Ross W, Turnbough CL, Jr, Gourse RL (1997) Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science* 278(5346):2092–2097.
18. Murray HD, Schneider DS, Gourse RL (2003) Control of rRNA expression by small molecules is dynamic and nonredundant. *Mol Cell* 12(1):125–134.
19. Barker MM, Gourse RL (2001) Regulation of rRNA transcription correlates with nucleoside triphosphate sensing. *J Bacteriol* 183(21):6315–6323.
20. Barne KA, Bown JA, Busby SW, Minchin SD (1997) Region 2.5 of the *Escherichia coli* RNA polymerase sigma 70 subunit is responsible for the recognition of the 'extended-10' motif at promoters. *EMBO J* 16(13):4034–4040.
21. Haugen SP, Ross W, Manrique M, Gourse RL (2008) Fine structure of the promoter-sigma region 1.2 interaction. *Proc Natl Acad Sci USA* 105(9):3292–3297.
22. Zhang Y, et al. (2012) Structural basis of transcription initiation. *Science* 349(6250):882–885.
23. Feklistov A, Darst SD (2011) Structural basis for promoter -10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell* 147(6):1257–1269.
24. Winkelman JT, et al. (2015) Crosslink mapping at amino acid-base resolution reveals the path of DNA in a scrunched initial transcribing complex. *Mol Cell* 59(5):768–780.
25. Chin JW, Martin AB, King DS, Wang L, Schultz PG (2002) Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc Natl Acad Sci USA* 99(17):11020–11024.
26. Vvedenskaya IO, et al. (2015) Massively systematic transcript readout (MASTER): Transcription start site selection, transcriptional slippage, and transcript yields. *Mol Cell* 60(6):953–965.
27. Winkelman JT, et al. (2016) Multiplexed protein-DNA crosslinking: Scrunching in transcription start site selection. *Science* 351(6277):1090–1093.
28. Gries TJ, Kontur WS, Capp MW, Saecker RM, Record MT, Jr (2010) One-step DNA melting in the RNA polymerase cleft opens the initiation bubble to form an unstable open complex. *Proc Natl Acad Sci USA* 107(23):10418–10423.

29. Ruff EF, et al. (2015) *E. coli* RNA polymerase determinants of open complex lifetime and structure. *J Mol Biol* 427(15):2435–2450.
30. Gourse RL (1988) Visualization and quantitative analysis of complex formation between *E. coli* RNA polymerase and an rRNA promoter in vitro. *Nucleic Acids Res* 16(20):9789–9809.
31. Paul BJ, Ross W, Gaal T, Gourse RL (2004) rRNA transcription in *Escherichia coli*. *Annu Rev Genet* 38:749–770.
32. Rao L, et al. (1994) Factor independent activation of rrnB P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. *J Mol Biol* 235(5):1421–1435.
33. Ross W, et al. (1993) A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* 262(5138):1407–1413.
34. Hirvonen CA, et al. (2001) Contributions of UP elements and the transcription factor FIS to expression from the seven *rrn P1* promoters in *Escherichia coli*. *J Bacteriol* 183(21):6305–6314.
35. Dickson RR, Gaal T, deBoer HA, deHaseth PL, Gourse RL (1989) Identification of promoter mutants defective in growth-rate-dependent regulation of rRNA transcription in *Escherichia coli*. *J Bacteriol* 171(9):4862–4870.
36. Ross W, Thompson JF, Newlands JT, Gourse RL (1990) *E. coli* Fis protein activates ribosomal RNA transcription in vitro and in vivo. *EMBO J* 9(11):3733–3742.
37. Brunner M, Bujard H (1987) Promoter recognition and promoter strength in the *Escherichia coli* system. *EMBO J* 6(10):3139–3144.
38. Burgess RR, Jendrisak JJ (1975) A procedure for the rapid, large-scale purification of *Escherichia coli* DNA-dependent RNA polymerase involving Polymin P precipitation and DNA-cellulose chromatography. *Biochemistry* 14(21):4634–4638.
39. Zaychikov E, et al. (1996) Mapping of catalytic residues in the RNA polymerase active center. *Science* 273(5271):107–109.
40. Artsimovitch I, Svetlov V, Murakami K, Landick R (2003) Co-overexpression of *Escherichia coli* RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J Biol Chem* 278(14): 12344–12355.
41. Svetlov V, Artsimovitch I (2015) Purification of bacterial RNA polymerase: Tools and protocols. *Methods Mol Biol* 1276:13–29.
42. Ross W, Schneider DA, Paul BJ, Mertens A, Gourse RL (2003) An intersubunit contact stimulating transcription initiation by *E. coli* RNA polymerase:interaction of the α C-terminal domain and σ region 4. *Genes Dev* 17(10):1293–1307.
43. Rutherford ST, Villers CL, Lee JH, Ross W, Gourse RL (2009) Allosteric control of *Escherichia coli* rRNA promoter complexes by DksA. *Genes Dev* 23(2):236–248.
44. Gosink KK, et al. (1993) DNA binding and bending are necessary but not sufficient for Fis-dependent activation of *rrnB* P1. *J Bacteriol* 175(6):1580–1589.