RNA polymerase gate loop guides the nontemplate DNA strand in transcription complexes

Monali NandyMazumdar[a,b], Yuri Nedialkov[a,b], Dmitri Svetlov[a,b,1], Anastasia Sevostyanova[a,b,2], Georgiy A. Belogurov[c], and Irina Artsimovitch[a,b,3]

[a]Department of Microbiology, The Ohio State University, Columbus, OH 43210; [b]The Center for RNA Biology, The Ohio State University, Columbus, OH 43210; and [c]Department of Biochemistry, University of Turku, Turku FIN-20014, Finland

Edited by Richard L. Gourse, University of Wisconsin-Madison, Madison, WI, and accepted by Editorial Board Member Kiyoshi Mizuuchi November 19, 2016 (received for review August 16, 2016)

Upon RNA polymerase (RNAP) binding to a promoter, the σ factor initiates DNA strand separation and captures the melted nontemplate DNA, whereas the core enzyme establishes interactions with the duplex DNA in front of the active site that stabilize initiation complexes and persist throughout elongation. Among many core RNAP elements that participate in these interactions, the β′ clamp domain plays the most prominent role. In this work, we investigate the role of the β gate loop, a conserved and essential structural element that lies across the DNA channel from the clamp, in transcription regulation. The gate loop was proposed to control DNA loading during initiation and to interact with NusG-like proteins to lock RNAP in a closed, processive state during elongation. We show that the removal of the gate loop has large effects on promoter complexes, trapping an unstable intermediate in which the RNAP contacts with the nontemplate strand discriminator region and the downstream duplex DNA are not yet fully established. We find that although RNAP lacking the gate loop displays moderate defects in pausing, transcript cleavage, and termination, it is fully responsive to the transcription elongation factor NusG. Together with the structural data, our results support a model in which the gate loop, acting in concert with initiation or elongation factors, guides the nontemplate DNA in transcription complexes, thereby modulating their regulatory properties.

RNA polymerase | transcription | discriminator | promoter | beta pincer

During each round of transcription, RNA polymerase (RNAP) establishes, maintains, and finally releases contacts with the DNA template and the RNA. These interactions mediate highly selective initiation, processive elongation, and precise termination and can be tuned to enable intricate regulation during the transcription cycle. RNAP resembles a crab claw in which the two pincers composed of the β′ and β subunits form an active site cleft that accommodates the nucleic acid chains (Fig. 1). The mobile β′ clamp domain that forms one of the pincers stands out as a central regulatory feature (1, 2). The open clamp likely allows the loading of the promoter DNA during initiation and the release of the template and the nascent RNA during termination. The clamp is thought to close to form an active initiation complex and to remain closed during elongation but may open partially at a hairpin-dependent pause site (3).

The clamp movements may be linked to those of the β lobe domain, which forms a part of the second pincer. The β gate loop (GL), which lies across the RNAP cleft from the tip of the clamp (Fig. 1), has been identified as an element that restricts the entry of the duplex promoter DNA into the narrow active site cleft (4), allowing only a single strand of DNA to pass through. This model posited that the DNA stands must separate outside the cleft before entry into RNAP, whereas footprinting studies demonstrated that the promoter DNA enters the active site cleft before it is opened (5). This controversy was a subject of an intense debate until a recent study revealed that the width of the RNAP cleft varies in solution and potentially is able to accommodate the duplex DNA (6).

As its gating role faded, a new role for the GL in transcript elongation has emerged. The GL was shown to interact with *Escherichia coli* RfaH (7), a specialized paralog of the essential transcription elongation factor NusG. Ubiquitous NusG homologs are thought to enhance elongation by bridging the βGL and the β′ clamp to stabilize the latter in a closed, pause-resistant conformation (8–10). Consistent with an important role of the GL, its sequence is relatively conserved in Bacteria (Fig. S1*A*), and it is essential for viability in *E. coli* (7). However, our findings that the enzyme lacking the GL did not exhibit defects in Rho-dependent termination (7), the essential function of NusG, suggested a different role for the GL.

In this work we found that, despite dramatic defects in growth (Fig. S1 *B* and *C*), the GL deletion conferred only mild defects in elongation, pausing, intrinsic termination, and RNA cleavage and did not abolish the response to the transcription elongation factors NusA and NusG in vitro. In contrast, the ΔGL enzyme exhibited strong defects during initiation: It formed unstable open complexes in which RNAP interactions with the downstream duplex DNA and the nontemplate (NT) DNA strand in the transcription bubble were compromised. Our results suggest that, as originally predicted (4), the GL guides the promoter DNA into its proper path in the RNAP holoenzyme.

Significance

The nontemplate DNA strand in the transcription bubble interacts with RNA polymerase and accessory factors to control initiation, elongation, transcription-coupled repair, and translation. During initiation, σ subunit interactions with the nontemplate DNA regulate promoter complex formation and lifetime, abortive synthesis, and start site selection. Here, we show that the β subunit gate loop contacts with an adjacent segment of the nontemplate discriminator region play a similar role during initiation. The deletion of the gate loop alters the structure and properties of promoter complexes and has pleiotropic effects on RNA chain elongation and termination. We propose that, acting in concert with accessory factors, the gate loop mediates the clamp closure and guides the nontemplate strand in initiation and elongation complexes.

Author contributions: M.N., Y.N., D.S., A.S., G.A.B., and I.A. designed research; M.N., Y.N., D.S., A.S., G.A.B., and I.A. performed research; M.N., Y.N., D.S., G.A.B., and I.A. analyzed data; and G.A.B. and I.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. R.L.G. is a Guest Editor invited by the Editorial Board.

[1]Present address: Department of Chemistry, University of Southern California, Los Angeles, CA 90007.

[2]Present address: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520.

[3]To whom correspondence should be addressed. Email: artsimovitch.1@osu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1613673114/-/DCSupplemental.

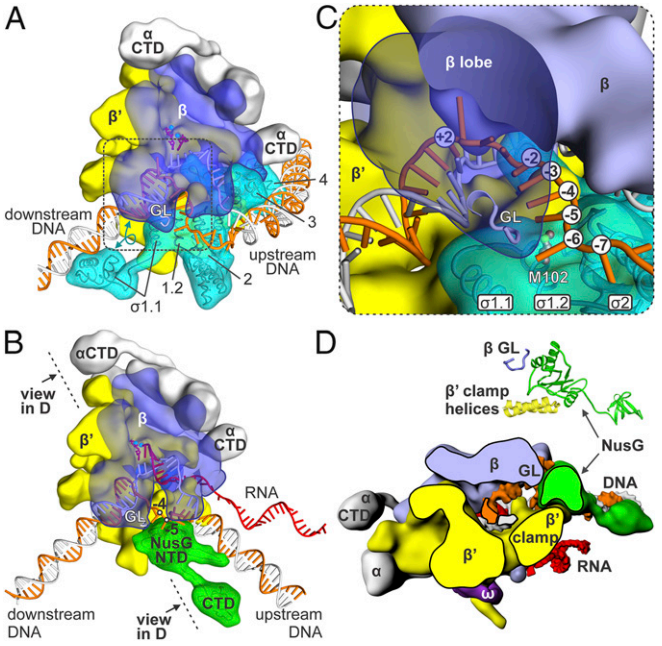

Fig. 1. Bacterial transcription complexes. (*A* and *B*) An overview of the open promoter complex (RPo) with a 6-nt discriminator (*A*) and the TEC with bound NusG (*B*). The composite models (Datasets S1 and S2) were generated using *T. thermophilus* RPo (22) and TEC (38) and elements from other structures as described in *SI Materials and Methods*. Proteins are depicted by simplified differentially colored molecular surfaces; β, σ^A and NusG are rendered semitransparent. The positions of the N-terminal domain (NTD) of the σ1.1 region, α C-terminal domains (CTDs), and NusG CTDs were chosen arbitrarily within the volume permitted by the length of the flexible linkers; the cyan arrow in *A* indicates that the NTD of σ1.1 is predominantly located near the β lobe domain in *E. coli* RPo (39). Nucleic acids and βGL are shown as cartoons, two Mg^{2+} ions in the active site are shown as cyan spheres, and an incoming NTP is shown as red sticks. Selected DNA nucleotides are numbered relative to the TSS in *A* and from the RNA 3′ end in the posttranslocated TEC in *B*. (*C*) A zoomed-in view of the GL and the discriminator region of the NT DNA (the rectangular area outlined in *A*). The β lobe and σA are semitransparent. *E. coli* σ^{70} Met102 was modeled into a homologous position of *T. thermophilus* σA and is shown in a balls-and-sticks configuration. (*D*) A side view of the TEC. Nucleic acids are depicted as surfaces. The view is clipped along the dashed line in *B* to expose NusG interactions with the GL and the β′ clamp domain. (*Inset*) NusG–RNAP contacts.

Results

Deletion of the GL Alters the Structure of Open Promoter Complexes. The discriminator region between the −10 hexamer and the transcription start site (TSS; +1) has recently emerged as an important modulator of the open complex properties (11–18). In a stable open complex formed by *Thermus thermophilus* RNAP, the −6 and −5 bases of the discriminator NT strand (NT^{DISC}) make crucial contacts with σ1.2, and βArg371 (*E. coli* numbering) interacts with the adjacent −4 and −3 bases (14). We hypothesized that the deletion of the GL may disrupt these interactions.

We first tested the effect of the GL at the bacteriophage λP_R promoter (Fig. 2*A*), which has been studied extensively by footprinting and kinetic approaches. Studies by Record and colleagues identified a series of λP_R promoter complexes that differ in RNAP–DNA interactions, including two unstable open intermediates, I_2 and I_3, and a very stable final open complex, RPo (19). During the isomerization from I_3 into RPo, RNAP tightens its interactions with NT^{DISC} and the downstream duplex DNA regions, massively stabilizing the complex (2, 19).

To determine whether the GL removal alters the NT DNA–RNAP interactions, we carried out footprinting with potassium permanganate, which modifies single-stranded or unstacked T residues. In the WT λP_R RPo, the NT strand T residues −10, −4, −3, and +2 were accessible to modification (Fig. 2*B*), consistent with previous studies (5). A different pattern was observed in the complex formed with the ΔGL RNAP: Although the −10 and +2 residues were as sensitive to modification as in the WT complex, the −4 and −3 residues were significantly protected (Fig. 2*B*). This reduced reactivity may indicate that, upon the loss of contacts with GL, these bases are inserted into a σ2 pocket, as observed with an unfavorable NT^{DISC} sequence (18), or are allowed to stack, as has been proposed for the I_2 intermediate at λP_R, in which a quantitatively similar pattern of reduced permanganate reactivity of the −4/−3 residues was observed with the WT RNAP (20).

RNAP contacts with the downstream DNA are established during the final steps of λP_R RPo formation (19) and are absent in earlier unstable open complexes such as I_2. We hypothesized that the downstream DNA interactions also may be destabilized in the ΔGL complex. To probe the latter contacts, we used exonuclease III footprinting (Fig. 2*C*). RPo formed by the WT RNAP protected the DNA from ExoIII digestion to +23, similar to the downstream boundary observed with other probes (19). In ΔGL complexes, the protection was significantly reduced, with Exo III able to digest DNA to +1 in a fraction of complexes. Together, the footprinting results suggest that the ΔGL open complex resembles the WT I_2 intermediate in which the NT DNA and the downstream duplex are not yet loaded into their tracks.

The GL Stabilizes Open Complexes. Substitutions of the discriminator bases and the σ and β residues that interact with the NT strand lead to decreases in open complex stability (11, 14, 16). To test whether the loss of the GL–NT^{DISC} contacts observed in the RPo structure (Fig. 1*D*) destabilizes the RPo, we carried out standard dissociation assays in which a preformed λP_R RPo is challenged with heparin, followed by measuring the remaining RPo by RNA synthesis (Fig. 3*A*). We found that deletion of the GL caused a 22-fold

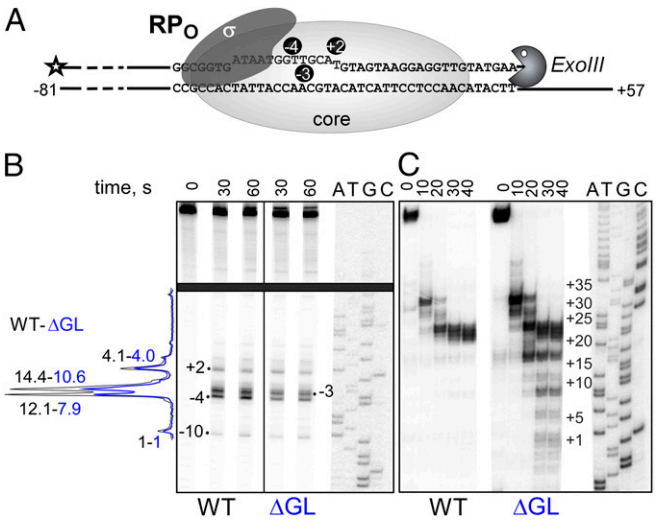

Fig. 2. Footprinting of the ΔGL open complexes. (*A*) A linear λP_R fragment in which the NT DNA was labeled at the 5′ end with [γ^{32}P]-ATP. (*B*) Preformed open complexes were treated with 2 mM $KMnO_4$ for 30 or 60 s. After quenching, ethanol precipitation, and piperidine cleavage, the DNA was analyzed on an 8% denaturing gel. A representative of three independent experiments is shown; the 0 point is an untreated DNA control. Traces of the 60-s reactions were generated with ImageQuant; the band intensities normalized to the −10 signal (taken as 1) are shown in black (WT) and blue (ΔGL). (*C*) ExoIII was added to preformed promoter complexes. Aliquots were quenched at the indicated times (0 represents an untreated DNA control) and analyzed on a 6% denaturing gel; a representative of three independent experiments is shown. The positions of the modified residues (*B*) or the protection boundary (*C*) were identified using sequencing ladders.

decrease in the λP_R RPo lifetime, from 65 to 3 min (Fig. 3*A*), comparable to the effects of substitutions of key NT bases and RNAP residues that interact with these bases (11, 14–16). Consistent with the deduction that T7A1 and *rrnB* P1 open complexes resemble λP_R I_3 and I_2 (19), the GL deletion had smaller (4.4- and 1.6-fold) effects at these promoters (Fig. 3*A*).

GL Contacts with the Discriminator. Permanganate footprinting (Fig. 2*B*) suggests that the removal of the GL alters RNAP interactions with the −4 and −3 bases at λP_R. To explore the conformational changes upon the loss of the GL further, we carried out cross-linking with a "zero-length" 4-thio-dT incorporated at the λP_R NT-4 position. This residue would be expected to cross-link to β and σ, as was indeed observed (Fig. 3*B*). With the WT RNAP, cross-linking to β was more efficient than to σ (~65 vs. 35%). When holoenzymes were assembled with σ carrying Y101A and M102A substitutions in σ1.2, which have been shown to weaken interactions with the adjacent NT^{DISC} residues (16), cross-linking to σ was diminished (~20–25%), as expected. With the ΔGL enzyme, cross-linking to the β and σ subunits was equally efficient (within the experimental error) (Fig. 3*B*), suggesting that the removal of the GL preferentially weakens the NT −4 base interactions with the β subunit. The residual cross-linking is likely caused by the repositioning of the NT^{DISC} region, e.g., toward β392 that has been shown to cross-link to D4 (21). Although the loss of cross-linking cannot be interpreted as evidence for direct GL–DNA contacts, these results are fully consistent with the GL–DNA contacts observed in crystal structures (14, 22, 23).

Interactions between the NT DNA and RNAP elements exhibit some sequence preference; e.g., C is strongly disfavored at D2, and G is preferred at +2 (14, 15). We wondered if GL–NT interactions are also sequence specific. We first tested the effect of simultaneously substituting the −3 and −4 T residues with C. This double substitution reduced open complex lifetime at λP_R by fourfold with the WT RNAP (Fig. 3*A*) but only by 1.5-fold with the ΔGL RNAP. Because both the GC content and purine/pyrimidine bases in the discriminator (18) alter open complex properties, we substituted −3 and −4 with A individually. These changes slightly destabilized the WT complexes but had an opposite effect on the ΔGL complexes (Fig. 3*A*). The effects were similar at both −3A and −4A promoters, reducing the GL contribution to ~13.5-fold.

The β residue Arg371 is highly conserved in Bacteria (Fig. S1*A*) and makes direct contacts with the −3 and −4 NT DNA residues in the bacterial RPo (14). To evaluate the contribution of Arg371 to the GL-specific effects on promoter complex properties, we replaced this residue with alanine and measured the stability of the open complexes formed on the WT and −3, −4 CC λP_R promoters. The βR371A substitution reduced the lifetime of the WT λP_R complexes approximately sevenfold (to 9.7 ± 0.9 min), compared with the 22-fold effect of removing the entire GL (Fig. 3*A*). On the CC promoter, the βR371A complexes were only slightly more stable (2.5 ± 0.2 min) than those formed by ΔGL RNAP (1.9 ± 0.3 min). These results suggest that Arg371 establishes functional, base-specific interactions with the discriminator region that stabilize the open promoter complexes.

Deletion of the GL Reduces Abortive Synthesis. *rrnB* P1 has an 8-nt discriminator with a C residue at the −7 (D2) position (Fig. 3*A*) that, together with a suboptimal spacer region, precludes productive contacts with σ1.2 (15). Despite forming very unstable complexes, *rrnB* P1 is one of the strongest promoters, in part

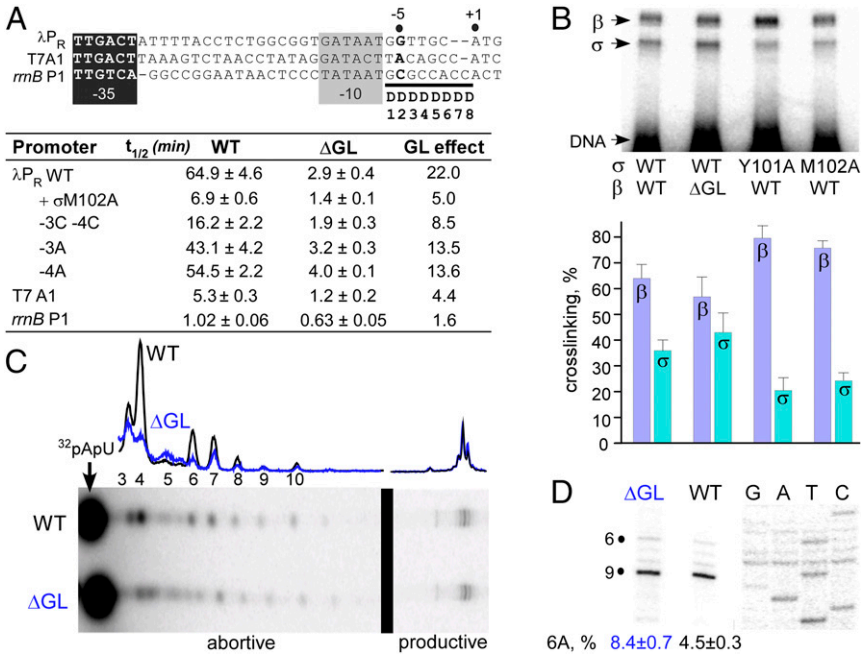

Fig. 3. Effects of the GL deletion on promoter complex properties. (*A*, *Upper*) Promoters used in this study. The −35 and −10 hexamers are boxed. The discriminator element is underlined, and its positions are numbered. The TSS and the D2 (−5 at λP_R) base that makes key contacts with σ1.2 are indicated. (*Lower*) The half-lives of open complexes (in minutes) were measured by assaying the fraction of transcriptionally competent complexes following the competitor challenge (*SI Materials and Methods*) in at least three independent repeats; data are shown as mean ± SD. The GL effect was defined as the ratio of the half-lives of the WT and ΔGL open complexes. (*B*, *Upper*) Open complexes assembled on λP_R with 4-thio-dT at the −4 NT position were exposed to 365-nm UV light. The reactions were separated on 4–12% Bis-Tris gels. (*Lower*) Relative cross-linking to σ and β (normalized to DNA) was calculated from four independent experiments; error bars indicate the SD. (*C*) Abortive initiation at λP_R. Open complexes were formed in the presence of [γ^{32}P]-ApU and were incubated with 250 μM NTPs (see Fig. S2). A representative gel and trace analysis of abortive products are shown. (*D*) TSS at *rrnB* P1 were mapped by primer extension of in vitro-transcribed RNAs. Extension products were analyzed on a denaturing 12% gel along with the sequencing ladder generated with the same primer. Positions corresponding to a TSS at 6 and 9 are indicated.

because it does not produce abortive RNAs. C–7G substitution restores favorable interactions with σ1.2, stabilizing the *rrnB* P1 promoter complex ~40-fold (15) and impeding promoter escape (17). A very stable λP_R RPo with a 6-nt discriminator makes favorable contacts with σ1.2 and produces abundant abortive products; the corresponding G–5C substitution reduces its half-life 14-fold (15). As is consistent with a stabilizing role of GL-NT^{DISC} contacts, we found that the GL deletion reduced abortive synthesis at λP_R (Fig. 3*C*). The major effect was observed with the 4-mer RNA, which is too short to clash directly with σ3.2 (24, 25), supporting a model in which a scrunched template strand stimulates RNA release (21).

The decrease in abortive synthesis by the ΔGL RNAP could be caused by reduced RPo stability (Fig. 3*A*) rather than by the loss of NT strand contacts. To evaluate this possibility, we tested the effects of mutations in RNAP that decreased or increased RPo stability (19, 26) on escape from λP_R. We found that neither destabilizing (β′ΔSI3 and β′R339A) nor stabilizing (σ^{70}Δ1–55) changes altered the pattern of abortive products (Fig. S2*A*). We conclude that decreasing the stability of the λP_R complex per se is not sufficient to facilitate promoter escape.

We next tested the effect of the discriminator–RNAP interactions on abortive synthesis. Increasing the distance between the −10 region and the start site is expected to weaken these contacts, facilitating promoter escape (17). In agreement with recent reports (12, 17), we found that the discriminator, but not other promoter regions, determined the differences in abortive synthesis between the T7A1 and λP_R promoters (Fig. S2*C*). The Y101A substitution in σ1.2, which compromises σ contacts with the NT^{DISC} (16), reduced abortive synthesis at λP_R and eliminated the RNAP ability to "sense" the length of the discriminator (Fig. S3).

Deletion of the GL Affects TSS Selection. RNAP can start transcription at a range of positions relative to the −10 hexamer (27), and σ1.2–NT^{DISC} interactions play a key role in TSS selection (17). WT *rrnB* P1 complexes, in which DNA is scrunched, initiate at 9A (+1), whereas the C–7G substitution or spacer insertions that reduce scrunching shift the TSS to a "standard" 6A (17). By altering the NT^{DISC} path through RNAP, the GL also may contribute to TSS choice. We found that the ΔGL RNAP used the 6A start site more efficiently than the WT enzyme (Fig. 3*D*), suggesting that the GL, which contacts the NT^{DISC} just upstream of the NT DNA segment extruded upon scrunching (18, 21), may stabilize the scrunched state.

The GL Is Largely Dispensable for Transcript Elongation Control. The location of the GL in the transcript elongation complex (TEC) suggests that it may affect RNAP processivity and response to NusG (Fig. 1 *A* and *C*), which inhibits pausing, particularly at sites where RNAP is prone to backtracking (28). We first tested the effect of the GL deletion on transcription through two tandem GGGAUGCGUGCG pause sites that fit the ubiquitous pause consensus (29). The pausing patterns for the WT and ΔGL enzymes were very similar (Fig. 4*A*). In the absence of NusG, the ΔGL RNAP reached the end of the template slightly sooner than the WT enzyme, whereas in the presence of NusG, the two enzymes elongated the nascent RNA at nearly the same rate (Fig. 4*B*). The modest defects in pausing suggest that the GL deletion may inhibit backtracking. This conclusion is supported by our observation that TECs formed with ΔGL enzyme were resistant to the nascent RNA cleavage stimulated by the transcript cleavage factor GreB (Fig. S4) and by observations of RNAP backtracking in real time (30).

The lack of a strong NusG effect contrasts with our earlier conclusion that the GL is required for RNAP modification by NusG (7). We identified an HTTT motif as a contact point of RfaH to the GL and replaced corresponding SWHL residues in NusG with four alanines. This substitution abolished the NusG antipausing activity but not the binding to the TEC because the same variant was fully capable of enhancing Rho-dependent termination (7). We therefore attributed the loss of the anti-pausing activity to the loss of NusG–GL contacts. However, the replaced residues likely also interact with the NT DNA (Fig. 4*C*). Crickard et al. recently reported that replacing six residues of Spt5, a yeast homolog of NusG, located on the same face as the SWHL NusG motif with alanines abolished Spt5 cross-linking to the NT DNA and its anti-arrest activity (31). Although it is uncertain if the altered surface residues (which are only weakly conserved) or more extensive structural changes (such as local refolding caused by altered core residues) lead to the observed phenotypes, it seems probable that the four-alanine substitution in NusG abolishes its anti-pausing activity by altering interactions with the NT DNA and not with the GL.

We next asked whether the removal of the GL may influence the clamp dynamics, altering RNAP response to hairpin-dependent sites such as *hisP*, at which interactions between the nascent RNA hairpin and the β flap domain induce opening of the clamp (3). Acting in concert with RfaH, which fills the gap between the GL and the clamp tip, the GL reduces RNAP pausing at *hisP* (7). However, consistent with our previous findings (7), deletion of the GL reduced pausing at the *hisP* site (Fig. S5*A*), arguing that the GL

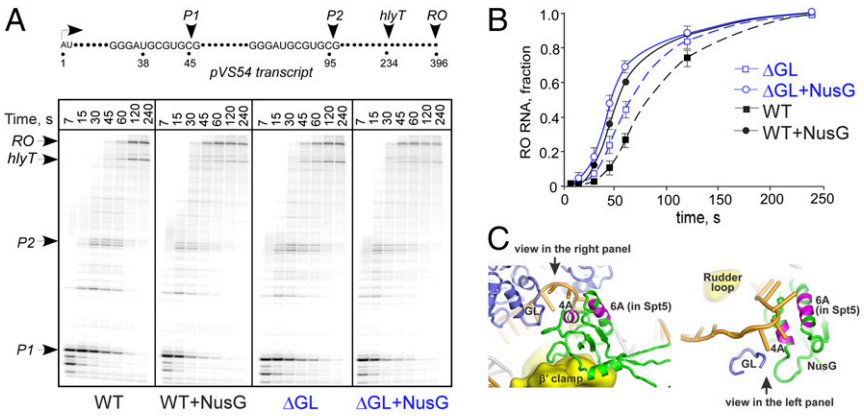

Fig. 4. The GL RNAP is responsive to NusG. (*A*) Single-round elongation assays were carried out on the pVS54 template; positions of pause sites, the *hlyT* terminator, and the run-off (RO) are shown. Halted [α^{32}P]-CMP–labeled A38 TECs formed with the WT or ΔGL RNAP were chased in the presence or absence of NusG. Reactions stopped at the indicated times were analyzed on 8% denaturing gels. (*B*) The average fraction (± SD) of run-off RNA determined from three independent experiments, including that shown in *A*. (*C*) Close-up view of NusG contacts with the GL and NT-DNA in the TEC model. The positions of residues replaced by alanine in NusG (7) and Spt5 (31) are shown in magenta.

does not favor the clamp closure in the absence of a bridging factor. Mild defects of the ΔGL RNAP observed at intrinsic terminators that use different mechanisms of release (Fig. S5*C*) are also likely caused by defects in pausing.

NusA, an essential transcription factor which acts in part through binding to an RNA hairpin as it emerges from the RNA exit channel (1), stimulates pausing and termination. A failure to respond to NusA could explain the growth defect of ΔGL RNAP. However, we found that NusA increased pausing and termination (Fig. S5) by the ΔGL and WT enzymes similarly. In summary, these and prior (7) results demonstrate that deletion of the GL leads to modest defects in elongation, transcript cleavage, and termination and does not abolish the RNAP response to general transcription factors, at least in vitro.

Discussion

In this work, we show that the β subunit GL modulates every step of the transcription cycle but has the most pronounced effects during initiation. The removal of the GL altered the structure and properties of promoter complexes, apparently blocking isomerization into a long-lived open complex that forms at some promoters, such as λP_R. We cannot exclude the possibility that the GL deletion alters the structure of transcription complexes indirectly, but several arguments support a hypothesis that all the effects of the GL are mediated through its interactions with the single-stranded (ss) NT DNA. First, the GL is a surface loop, which we replaced with two glycine residues to preserve the positions of the flanking β regions. Second, the GL residue Arg371 makes direct contacts with the NT DNA in promoter complex structures (14, 22, 23), and our results show that the βR371A substitution destabilizes promoter complexes. Third, the GL does not interact with other core RNAP regions. Fourth, although the GL closely approaches σ1.2, these hypothetical contacts cannot (*i*) explain the GL effects during elongation or (*ii*) fully account for its effects on open complexes, because a large part of the GL effect depends on the discriminator sequence, and a "destabilizing" M102A substitution in σ1.2 and a longer discriminator region in T7A1 reduce the GL effect similarly (Fig. 3*A*).

During initiation, the GL forms part of a relay in the NT strand interactions that commence with σ capturing the −11A base in a pocket, unzipping the −10 region to place the −7 base in another pocket (32), and making contacts with two bases downstream of the −10 hexamer (14, 16). The DNA then bends into the channel, with the GL stabilizing the NT strand and guiding it toward the β CRE pocket, which may capture the +2 base in a stable final open complex (14). Once RNAP escapes from the promoter, the GL may continue to guide the NT strand, sometimes acting in concert with elongation factors that take the place of the released σ.

GL–NT DNA Interactions in Initiation. Bacterial promoters exhibit astonishing diversity in sequences, strength, and sensitivity to regulation. Although some, such as *rrnB* P1, are exquisitely tunable by cellular cues, others, such as λP_R, appear to be optimized for steady RNA output. These properties are determined, in part, by the differing structures of open complexes. Record and colleagues argued that open complex intermediates transiently populated at λP_R correspond to final open complexes at other promoters, with λP_R I_2 resembling the *rrnB* P1 open complex (19). Footprinting analysis shows that in I_2 the contacts with NT DNA are loose, the clamp is not locked, and the downstream DNA is not held tightly (19, 20). Our results support a model in which RNAP interactions with the discriminator mediate the transition from I_2 to RPo and implicate the GL in this isomerization step.

The discriminator region was proposed to direct an elaborate cascade of interactions and conformational changes that determine the structure of open complexes at different promoters (19). At λP_R, a short 6-nt discriminator with G at D2 favors tight interactions with RNAP and enables dramatic, 10^5-fold stabilization of the final RPo relative to an early, competitor-sensitive I_2. At T7A1, a longer discriminator with a suboptimal A at D2 weakens the contacts, reducing the extent of stabilization to ∼250-fold and mimicking λP_R I_3. Finally, *rrnB* P1 appears never to progress beyond the initial (I_2-like) unstable open complex because its 8-nt discriminator with C at D2 cannot form productive contacts with RNAP. In this model, RNAP elements that interact with the discriminator are expected to be critically important at promoters that form long-lived open complexes but largely dispensable at promoters forming unstable complexes. Indeed, alanine substitution of σM102, which makes a van der Waals contact with G at D2 (14, 16), dramatically destabilizes very stable complexes formed at λP_R and *rrnB* P1 C–7G variant (16) but has only a small effect at WT *rrnB* P1 (16). Similarly, deletion of the GL has a 22-fold effect at λP_R, compared with 4.4- and 1.6-fold effects at T7A1 and *rrnB* P1, respectively (Fig. 3*A*).

It is also possible that ternary interactions between GL, NT^{DISC}, and σ1.2 not only position the bubble and keep it from collapsing but also stabilize the clamp in a closed conformation. Notably, promoter complex intermediates have been proposed to have an open clamp conformation (19).

GL–NT DNA Interactions in Elongation. In contrast to its well-established role in initiation, the NT DNA effects on elongation are less clear because of the lack of experimental evidence on the NT DNA path in TEC. Analysis of RNAP lacking the GL could provide insights into the role of the NT strand in elongation. The GL is located nearly 60 Å from the RNAP active site and more than 20 Å from the upstream and downstream DNA duplexes, but deletion of the GL reduces RNAP backtracking (Fig. 4*A* and Fig. S4) and pausing at *his*P (Fig. S5*A*). We speculate that removal of the GL alters the path of the ss NT DNA through the TEC. The NT DNA is absent or unresolved in TEC structures but is well resolved in factor-stabilized initiation and initially transcribing complexes (14, 22, 33, 34). In the latter structures, the NT DNA separates from the template DNA immediately downstream of the β fork loop, passes along the inner side of the GL, exits the cleft between the β lobe and β protrusion domains, loops around σ2, and rejoins the template DNA to form the upstream DNA duplex (Fig. 1). The NT DNA can be modeled to follow a similar path in the TEC (Fig. 1*A*), looping around the NusG NTD instead of σ2. These contacts likely account for sequence-specific interactions with NusG homologs (29, 35, 36). In this model, the GL restricts the downstream portion of the bubble inside the cleft, biasing the NT DNA to loop out upstream and facilitating DNA reannealing downstream (Fig. S6), thereby promoting backtracking and pausing at *his*P, which occurs in a pretranslocated register. In the absence of the GL, the NT DNA could adopt an unconstrained conformation (Fig. S6), eliminating the backward translocation bias and reducing RNAP sensitivity to backtrack-prone and pretranslocated pauses.

The NT DNA has been implicated in the action of accessory factors that target the elongating RNAP (29, 35–37). It is highly likely that, as in σ, these regulators establish simultaneous contacts with the GL and NT DNA. Although some factors (e.g., RfaH) critically depend on both sets of contacts, others (e.g., *E. coli* NusG) do not (Fig. 4*A*). We stand by our hypothesis that the protein–protein contacts of RfaH and NusG with the GL restrict the clamp (7), but we now suggest that these paralogous factors lock the clamp in different states. Although NusG may stabilize the clamp in a relatively open conformation (10), the smaller RfaH may restrict the clamp in a more closed state. These differences would explain why RfaH reduces pausing at the *hisP* site that is accompanied by the clamp opening (3), whereas NusG does not.

In conclusion, our results and the available structural data support a model in which the GL influences the regulatory properties of transcription complexes through a combination of direct contacts with the NT strand and with the initiation/elongation factors.

Materials and Methods

Details for all procedures are in *SI Materials and Methods*. Plasmids are listed in Table S1, and oligonucleotides are listed in Table S2.

Elongation Assays. Halted [α^{32}P]-NMP–labeled TECs were formed on linear DNA templates with ApU dimer and starting NTP subsets. Following incubation with NusG (100 nM), where indicated, transcription was restarted by the addition of nucleotides at the concentrations indicated in the figures and 50 μg/mL rifapentine at 37 °C to limit the elongation to a single round. Aliquots were withdrawn at selected times, quenched, and analyzed on denaturing urea-acrylamide gels.

Open Complex Stability Assays. Linear DNA templates were incubated with RNAP holoenzyme at 37 °C for 15 min. At time 0, a competitor was added (heparin at 20 μg/mL for λP_R, heparin at 10 μg/mL for T7A1, or a 200 nM consensus promoter DNA fragment for *rrnB* P1). Aliquots were withdrawn at selected times and were added to a prewarmed mixture of a dinucleotide primer and NTP substrates specified by the promoter sequence, including an [α^{32}P]-NTP.

Footprinting Analysis. The linear λP_R promoter fragment was made by PCR amplification with the [γ^{32}P]-ATP–labeled NT strand primer 17. Open complexes were assembled with the WT or ΔGL RNAP and were probed with $KMnO_4$ or ExoIII (New England Biolabs). The positions of modified/protected residues were identified using sequencing ladders generated with the same primer.

ACKNOWLEDGMENTS. We thank Richard L. Gourse and Tomacz Heyduk for comments on the manuscript, Emily Ruff and Tom Record for numerous discussions, and Tom Santangelo for a gift of pTS111.

1. Belogurov GA, Artsimovitch I (2015) Regulation of transcript elongation. *Annu Rev Microbiol* 69:49–69.
2. Ruff EF, Record MT, Jr, Artsimovitch I (2015) Initial events in bacterial transcription initiation. *Biomolecules* 5(2):1035–1062.
3. Hein PP, et al. (2014) RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat Struct Mol Biol* 21(9):794–802.
4. Vassylyev DG, et al. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature* 417(6890):712–719.
5. Davis CA, Bingman CA, Landick R, Record MT, Jr, Saecker RM (2007) Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by Escherichia coli RNA polymerase. *Proc Natl Acad Sci USA* 104(19):7833–7838.
6. Chakraborty A, et al. (2012) Opening and closing of the bacterial RNA polymerase clamp. *Science* 337(6094):591–595.
7. Sevostyanova A, Belogurov GA, Mooney RA, Landick R, Artsimovitch I (2011) The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol Cell* 43(2):253–262.
8. Klein BJ, et al. (2011) RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc Natl Acad Sci USA* 108(2):546–550.
9. Martinez-Rucobo FW, Sainsbury S, Cheung AC, Cramer P (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J* 30(7):1302–1310.
10. Schulz S, et al. (2016) TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc Natl Acad Sci USA* 113(13):E1816–E1825.
11. Petushkov I, Pupov D, Bass I, Kulbachinskiy A (2015) Mutations in the CRE pocket of bacterial RNA polymerase affect multiple steps of transcription. *Nucleic Acids Res* 43(12):5798–5809.
12. Vvedenskaya IO, et al. (2014) Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science* 344(6189):1285–1289.
13. Vvedenskaya IO, et al. (2016) Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection. *Proc Natl Acad Sci USA* 113(21):E2899–E2905.
14. Zhang Y, et al. (2012) Structural basis of transcription initiation. *Science* 338(6110): 1076–1080.
15. 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.
16. 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.
17. Winkelman JT, Chandrangsu P, Ross W, Gourse RL (2016) Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of E. coli ribosomal RNA promoters. *Proc Natl Acad Sci USA* 113(13):E1787–E1795.
18. Winkelman JT, et al. (2016) Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. *Science* 351(6277):1090–1093.
19. Ruff EF, et al. (2015) E. coli RNA polymerase determinants of open complex lifetime and structure. *J Mol Biol* 427(15):2435–2450.
20. 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.
21. Winkelman JT, et al. (2015) Crosslink mapping at amino acid-base resolution reveals the path of scrunched DNA in initial transcribing complexes. *Mol Cell* 59(5):768–780.
22. Bae B, Feklistov A, Lass-Napiorkowska A, Landick R, Darst SA (2015) Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *eLife* 4:4.
23. Zuo Y, Steitz TA (2015) Crystal structures of the E. coli transcription initiation complexes with a complete bubble. *Mol Cell* 58(3):534–540.
24. Sevostyanova A, Artsimovitch I (2010) Functional analysis of Thermus thermophilus transcription factor NusG. *Nucleic Acids Res* 38(21):7432–7445.
25. Andrecka J, et al. (2009) Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. *Nucleic Acids Res* 37(17):5803–5809.
26. Pupov D, et al. (2010) Multiple roles of the RNA polymerase beta' SW2 region in transcription initiation, promoter escape, and RNA elongation. *Nucleic Acids Res* 38(17):5784–5796.
27. Vvedenskaya IO, et al. (2015) Massively systematic transcript end readout, "MASTER": Transcription start site selection, transcriptional slippage, and transcript yields. *Mol Cell* 60(6):953–965.
28. Artsimovitch I, Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci USA* 97(13):7090–7095.
29. Larson MH, et al. (2014) A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* 344(6187):1042–1047.
30. Turtola M, Belogurov GA (2016) NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *eLife* 5:18096.
31. Crickard JB, Fu J, Reese JC (2016) Biochemical analysis of yeast suppressor of Ty 4/5 (Spt4/5) reveals the importance of nucleic acid interactions in the prevention of RNA polymerase II arrest. *J Biol Chem* 291(19):9853–9870.
32. Feklistov A, Darst SA (2011) Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. *Cell* 147(6):1257–1269.
33. Barnes CO, et al. (2015) Crystal structure of a transcribing RNA polymerase II complex reveals a complete transcription bubble. *Mol Cell* 59(2):258–269.
34. Basu RS, et al. (2014) Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J Biol Chem* 289(35):24549–24559.
35. Belogurov GA, et al. (2007) Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol Cell* 26(1):117–129.
36. Yakhnin AV, Murakami KS, Babitzke P (2016) NusG is a sequence-specific RNA polymerase pause factor that binds to the non-template DNA within the paused transcription bubble. *J Biol Chem* 291(10):5299–5308.
37. Epshtein V, et al. (2014) UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* 505(7483):372–377.
38. Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* 448(7150):157–162.
39. Mekler V, et al. (2002) Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* 108(5):599–614.
40. Svetlov V, Artsimovitch I (2015) Purification of bacterial RNA polymerase: Tools and protocols. *Methods Mol Biol* 1276:13–29.
41. Belogurov GA, Sevostyanova A, Svetlov V, Artsimovitch I (2010) Functional regions of the N-terminal domain of the antiterminator RfaH. *Mol Microbiol* 76(2):286–301.
42. Bernecky C, Herzog F, Baumeister W, Plitzko JM, Cramer P (2016) Structure of transcribing mammalian RNA polymerase II. *Nature* 529(7587):551–554.
43. Opalka N, et al. (2010) Complete structural model of Escherichia coli RNA polymerase from a hybrid approach. *PLoS Biol* 8(9):e1000483.
44. Fiser A, Sali A (2003) ModLoop: Automated modeling of loops in protein structures. *Bioinformatics* 19(18):2500–2501.
45. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12–21.
46. Ederth J, Artsimovitch I, Isaksson LA, Landick R (2002) The downstream DNA jaw of bacterial RNA polymerase facilitates both transcriptional initiation and pausing. *J Biol Chem* 277(40):37456–37463.
47. Blaby-Haas CE, Furman R, Rodionov DA, Artsimovitch I, de Crécy-Lagard V (2011) Role of a Zn-independent DksA in Zn homeostasis and stringent response. *Mol Microbiol* 79(3): 700–715.
48. Elgamal S, Artsimovitch I, Ibba M (2016) Maintenance of transcription-translation coupling by elongation factor P. *MBio* 7(5):e01373–16.
49. Santangelo TJ, Roberts JW (2004) Forward translocation is the natural pathway of RNA release at an intrinsic terminator. *Mol Cell* 14(1):117–126.
50. Belogurov GA, Mooney RA, Svetlov V, Landick R, Artsimovitch I (2009) Functional specialization of transcription elongation factors. *EMBO J* 28(2):112–122.
51. Artsimovitch I, Landick R (2002) The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* 109(2):193–203.
52. Larson MH, Greenleaf WJ, Landick R, Block SM (2008) Applied force reveals mechanistic and energetic details of transcription termination. *Cell* 132(6):971–982.