NusG Is a Sequence-specific RNA Polymerase Pause Factor That Binds to the Non-template DNA within the Paused Transcription Bubble*

Received for publication, November 13, 2015, and in revised form, January 6, 2016 Published, JBC Papers in Press, January 7, 2016, DOI 10.1074/jbc.M115.704189

Alexander V. Yakhnin, Katsuhiko S. Murakami, and Paul Babitzke

From the Department of Biochemistry and Molecular Biology, Center for RNA Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802

NusG, referred to as Spt5 in archaeal and eukaryotic organisms, is the only transcription factor conserved in all three domains of life. This general transcription elongation factor binds to RNA polymerase (RNAP) soon after transcription initiation and dissociation of the RNA polymerase σ factor. Escherichia coli NusG increases transcription processivity by suppressing RNAP pausing, whereas Bacillus subtilis NusG dramatically stimulates pausing at two sites in the untranslated leader of the trpEDCFBA operon. These two regulatory pause sites participate in transcription attenuation and translational control mechanisms, respectively. Here we report that B. subtilis NusG makes sequence-specific contacts with a T-rich sequence in the non-template DNA (ntDNA) strand within the paused transcription bubble. NusG protects T residues of the recognition sequence from permanganate oxidation, and these T residues increase the affinity of NusG to the elongation complex. Binding of NusG to RNAP does not require interaction with RNA. These results indicate that bound NusG prevents forward movement of RNA polymerase by simultaneously contacting RNAP and the ntDNA strand. Mutational studies indicate that amino acid residues of two short regions within the NusG N-terminal domain are primarily responsible for recognition of the trp operon pause signals. Structural modeling indicates that these two regions are adjacent to each another in the protein. We propose that recognition of specific sequences in the ntDNA and stimulation of RNAP pausing is a conserved function of NusG-like transcription factors.

Although transcription elongation by RNAP² is highly progressive, it is punctuated by random and programmed pausing events. Transcription elongation factors, such as NusA and NusG, can dramatically influence the half-life of paused elongation complexes (PEC). Hairpin-stimulated pausing has been extensively studied in the untranslated leader of the Escherichia coli his operon. The multipartite his pause signal includes a pause RNA hairpin, the sequence between the hairpin and RNA 3' end, and the downstream DNA sequence (1, 2). Pausing at the his site is stimulated by NusA but is unaffected by NusG (3). NusA stimulates interaction of the pause hairpin with the β-flap domain of RNAP that favors a catalytically inactive conformation of the enzyme (4). E. coli contains a NusG paralog called RfaH, which stimulates hairpin-independent pausing at ops sites. RfaH binds specifically to the ops sequence within the ntDNA strand of the transcription bubble (3, 5, 6). RfaH modifies RNAP into a form that is resistant to Rho-dependent termination and also assists in co-transcriptional translation initiation of genes with weak Shine-Dalgarno sequences via direct interaction with the 30S ribosomal subunit. RfaH has the strongest stimulatory effect on expression of horizontally transferred genes that have high A+T content (7, 8).

The Bacillus subtilis trpEDCFBA operon contains six of seven genes for de novo tryptophan biosynthesis. Expression of the trp operon is negatively regulated by transcription attenuation and translation repression mechanisms in response to the availability of tryptophan. The trp RNA-binding attenuation protein (TRAP) is both the sensor of the intracellular tryptophan level and the regulator of trp operon expression (reviewed in Ref. 9). The presence of overlapping antiterminator and intrinsic terminator structures is a key feature of the attenuation mechanism. Tryptophan-activated TRAP binds to the untranslated leader of the nascent mRNA and prevents formation of the antiterminator structure. As a consequence, the downstream intrinsic terminator forms and transcription stops upstream of the trp operon structural genes. In limiting tryptophan conditions, TRAP does not bind to the trp transcript, the antiterminator forms, and the operon is expressed (10). In the translation control mechanism, TRAP binding to trp operon readthrough transcripts promotes formation of the trpE Shine-Dalgarno-sequestering hairpin, thereby inhibiting translation by preventing ribosome binding. In the absence of bound TRAP, an alternative RNA structure forms such that the translation initiation region is single-stranded and available for ribosome binding (11, 12).

NusA and NusG cooperatively stimulate RNAP pausing at two pause sites (U107 and U144) within the B. subtilis trp leader. Pausing is capable of synchronizing the position of RNAP with TRAP binding. As U107 is just upstream of the trp operon structural genes. In limiting tryptophan conditions, TRAP does not bind to the trp transcript, the antiterminator forms, and the operon is expressed (10). In the translation control mechanism, TRAP binding to trp operon readthrough transcripts promotes formation of the trpE Shine-Dalgarno-sequestering hairpin, thereby inhibiting translation by preventing ribosome binding. In the absence of bound TRAP, an alternative RNA structure forms such that the translation initiation region is single-stranded and available for ribosome binding (11, 12).

NusA and NusG cooperatively stimulate RNAP pausing at two pause sites (U107 and U144) within the B. subtilis trp leader. Pausing is capable of synchronizing the position of RNAP with TRAP binding. As U107 is just upstream of the critical overlap between the antiterminator and terminator structures, pausing at this position provides additional time for TRAP to bind and promote termination (13). Pausing at U144
NusG-stimulated Pausing

participates in the translation repression mechanism by providing a second opportunity for TRAP to bind to the nascent trp leader transcript (14). NusG has a much greater stimulatory effect on RNAP pausing at the U144 pause site than NusA. Moreover, NusG is required for pausing at the U144 pause site in vivo and proper regulation of the trp operon (15).

The pause-stimulating activity of B. subtilis NusG is in sharp contrast to the pause-suppressing activity of E. coli NusG (3, 16). Of particular interest, E. coli NusG does not influence B. subtilis RNAP pausing at U144 but competitively inhibits B. subtilis NusG-stimulated pausing when both factors are present together in vitro (15). This difference between two otherwise similar NusG proteins raises the question as to what specific amino acids are responsible for two fundamentally distinct effects on pausing. We conducted a systematic mutagenesis of B. subtilis NusG by substituting single amino acids and larger regions with the corresponding sequences from E. coli NusG and identified two regions that are critical for NusG-stimulated pausing. We also found that B. subtilis NusG responds to the U144 pause signal by sequence-specific recognition of the ntDNA strand within the paused transcription bubble.

Experimental Procedures

DNA Templates and Proteins—The untranslated leader of the B. subtilis trp operon contains a cryptic promoter (14). To avoid problems associated with transcription initiation from two promoters, the template for in vitro transcription was obtained by PCR amplification of the untranslated leader region of the B. subtilis trp operon (−1 to +175 relative to the natural start of transcription). During PCR amplification, the cryptic promoter was converted into a consensus promoter with an extended −10 element. This synthetic promoter directed transcription initiation at residue +37 in the trp leader region. PCR fragments were purified with the QIAquick PCR purification kit (Qiagen) and used as templates for in vitro transcription. A vector expressing the B. subtilis NusG N-terminal (NGN) domain was generated by deletion of the C-terminal sequence encoding the Kyprides-Onzoniós-Woeser (KOW) domain. Mutations in NGN were generated according to the QuikChange protocol (Strategene). His-tagged NGN proteins were overproduced and purified as described previously for NusG (15). His-tagged B. subtilis RNAP was purified as described (14, 17).

In Vitro Pausing of Promoter-initiated Transcription—Single-round in vitro transcription reactions and data analysis were performed as described previously (13, 18). Stable transcription elongation complexes containing a 29-nt-long transcript were formed in reaction mixtures containing ATP and GTP (8 μM each), 2 μM UTP, and 1 μCi/μl [α-32P]UTP at 37 °C. Transcription elongation was halted at position 65 of the trp leader due to the absence of CTP. Elongation of halted transcription complexes was resumed by the addition of all four NTPs (150 μM final concentration) together with 100 μg/ml heparin at 23 °C. When used, NusG or NGN proteins were added at 1 μM in the second step of the reaction. Aliquots of the elongation reaction were removed at various times. The last aliquot was chased for an additional 10 min at 37 °C with 0.5 mM of each NTP. Reactions were stopped by mixing with gel loading solution (95% formamide, 20 mM EDTA, pH 8.0, 0.2% SDS, 0.3 mg/ml bromophenol blue, and 3 mg/ml xylene cyanol), and samples were fractionated through standard 5% sequencing gels. Pause half-life and pause efficiency values were calculated by plotting relative intensities of the U144 pause band against the incubation time and fitting the data with a single exponential equation as described previously (13, 19).

Permanganate Footprinting—In vitro KMnO4 footprinting of promoter-initiated single-round in vitro transcription reactions followed published procedures with modifications (14, 15). DNA templates were labeled at the 5′ end of the ntDNA strand. Three-μl aliquots of in vitro transcription reactions were collected at various times after transcription of the halted elongation complex was resumed. An equal volume of 4 mM KMnO4 was then added, and incubation was continued for 1 min at 23 °C. Permanganate-mediated oxidation was stopped by the addition of 5 volumes of stop/cleavage solution (10 mM EDTA, pH 8.0, 40 mM β-mercaptoethanol, 0.15 mg/ml calf thymus DNA, and 7% (v/v) piperidine). Modified DNA was cleaved by incubation for 15 min at 90 °C. Piperidine was then removed by evaporation. DNA pellets were washed with acetone and dried, followed by two cycles of distilled H2O and evaporation. Dried pellets were dissolved in gel loading solution, and samples were fractionated through standard 5% sequencing gels.

PEC Assembly Using Nucleic Acid Scaffolds—Nucleic acid scaffolds for reconstitution of PECs were assembled as described previously with modifications (20, 21). For scaffolds in which the ntDNA and tDNA strands were not complementary within the transcription bubble, annealing reaction mixtures (3 μl) containing 1.5 μM template DNA (tDNA), 1.5 μM RNA, 3 μM ntDNA, and 0.15 mg/ml acetylated bovine serum albumin in 1.5× transcription buffer (1× transcription buffer is 40 mM Tris-HCl, pH 8.0, 5 mM MgCl2, 4 mM DTT, and 5% trehalose) were incubated at 37 °C for 10 min. The upstream tDNA was not complementary to the RNA to prevent an extended RNA-DNA hybrid. Reconstitution of PECs was performed by the addition of 1 μl of 3 μM RNAP in 1× transcription buffer to nucleic acid scaffolds, and incubation was continued at 37 °C for 10 min. This nucleic acid scaffold positioned RNAP at nucleotide 143 of the trp leader RNA. RNAs were 3′-end-labeled and simultaneously walked to the U144 pause site by incorporation of [α-32P]UTP (1.5 μl, 15 μCi, 3000 Cu/ml) for 1 min at room temperature. PECs were transferred to ice and stored up to 2 h before use. When fully complementary ntDNA and tDNA strands were used, RNA and tDNA were pre-annealed and incubated with RNAP at 37 °C for 10 min followed by incubation with ntDNA for an additional 10 min at 37 °C.

For crosslinking experiments, PECs were assembled with 5′-end-labeled ntDNA with 5-bromo-2′-deoxyuracil (5BrU) substituted at position T135 or T138. In this case, the RNA-labeling step was omitted. PECs were supplemented with NaCl, heparin, and Tween 20 to final concentrations of 25 mM, 50 μg/ml, and 0.1%, respectively. Ten-μl PEC aliquots were transferred to a 96-well microtiter plate and placed on ice. The plate was illuminated with a 312 nm UV lamp with a distance of 2 cm for 10 min. Crosslinked reactions were concentrated by evaporation and fractionated in 10% (20:1 acrylamide: bisacrylamide ratio) sequencing gels.
FIGURE 1. NusG protects T137–T139 within the U144 PEC from permanganate oxidation. A, sequence of the ntDNA, tDNA, and RNA strands within and adjacent to the U144 PEC derived from promoter-initiated in vitro transcription. Numbering corresponds to the U144 pause site. Positions of T135C and T137C mutations are shown. . . indicates that upstream and downstream sequences are not shown, including the native pause RNA hairpin. B, permanganate footprinting of the ntDNA strand of the U144 PEC was carried out by performing promoter-initiated single-round in vitro transcription reactions using templates labeled at the 5′ end of the ntDNA strand. Reactions were collected at the times indicated above each lane, and the ntDNA strand was then probed for oxidation by potassium permanganate and can be mapped by polyacrylamide gel electrophoresis.

In Vitro Pause Escape Assays on Assembled PECs—PECs containing 3′ end-labeled RNA were diluted in transcription buffer to desired volumes (50–100 nM PEC). Heparin and KCl were added to final concentrations of 100 µg/ml and 80 mM, respectively. NusG and/or NusA were then added to a final concentration of 1.7 µM. Extension was performed at 23 °C by mixing 3 volumes of PEC and 2 volumes of NTPs in transcription buffer. The identities of added NTPs and one 3′-dNTP were adjusted to sequences specified by the tDNA to prevent extension beyond a few nt. PECs reconstituted with Tcr3 tDNA plus RNA 20 and T2–11 tDNA plus RNA 11 were elongated to A147 with an extension mix containing 5 µM GTP, 5 µM UTP, and 50 µM 3′-dATP (see Table 1). PECs reconstituted with T2 tDNA plus RNA 35 were elongated to G148 with an extension mix containing 25 µM ATP, 25 µM UTP, and 50 µM 3′-dGTP (see Table 1). Samples were removed at various times, quenched with an equal volume of formamide loading dye, and analyzed by denaturing 12.5% (90:1 acrylamide:bisacrylamide ratio) polyacrylamide gel electrophoresis.

Results

NusG Makes Sequence-specific Contacts with the ntDNA Strand within the PEC—There are several unpaired T residues in the ntDNA strand within the U144 paused transcription bubble (Fig. 1A). These single-stranded T residues are sensitive to oxidation by potassium permanganate and can be mapped by permanganate footprinting. In a prior study, we demonstrated that the T residues between positions T134 and T139 are more reactive than the T residues between T141 and T144 (15). Our previously published results also demonstrated that NusG protects residues T137, T138, and T139 from permanganate oxidation with a concomitant increase in permanganate reactivity of residues T141 and T142 (15). In addition, mutations that altered residues T135, T137, T138, and/or T139 reduced or eliminated NusG-stimulated pausing (18). Thus, footprinting experiments were performed with WT, T135C, and T137C templates to determine whether mutations in this region altered the pattern of NusG-dependent permanganate reactivity. As we previously observed, NusG increased the duration of the pause and protected T137–T139 from permanganate oxidation (Fig. 1B). Moreover, NusG-dependent protection of these residues was lost in the T135C and T137C mutant templates. The permanganate reactivity pattern of the −NusG samples for the mutant templates was qualitatively indistinguishable from the corresponding +NusG samples (data not shown). Except for the complete loss of bands corresponding to the mutated T residues and a longer duration of the footprint, the permanganate reactivity of the mutant templates in the presence of NusG was essentially identical to the WT template in the absence of this protein (Fig. 1B). These results suggested that NusG makes sequence-specific contacts with the ntDNA strand of the U144 paused transcription bubble.

NusG consists of an NGN domain that interacts with the clamp helices of the RNAP β′ subunit and a C-terminal KOW domain that interacts with other factors, including termination factor Rho and ribosomes. The E. coli NGN domain (residues 1–123) is sufficient to bind RNAP, suppress pausing, and enhance transcript elongation in vitro (7, 22). We tested whether the B. subtilis NGN domain (residues 1–118) is sufficient for its pause-stimulating activity at U144 by performing promoter-initiated in vitro transcription reactions. Both full-length NusG and the NGN domain produced indistinguishable pausing patterns with similar pause half-lives and efficiencies (Fig. 2A). Therefore, the NGN domain alone is sufficient for modulation of transcription elongation by NusG. This observation allowed us to use the B. subtilis NGN domain in most of our subsequent pausing experiments.

We reconstituted active transcription elongation complexes using a variety of scaffolds consisting of DNA and RNA oligonucleotides. The first scaffold that we used was based on a published scaffold suitable for determination of the crystal structure of the eukaryotic Pol II elongation complex in which the RNA–DNA hybrid was restricted to 9 bp (23). This scaffold was generated by hybridization of partially complementary tDNA, ntDNA, and RNA oligonucleotides (TCr3, NT2, and RNA 20, respectively, Table 1), followed by RNAP binding (Fig. 2B). As the RNA ended 1 nt short of the pause site (i.e. position C143), we walked RNAP to...
U144 by the addition of [α-32P]UTP, thereby generating PECs labeled at the RNA 3′ end (Fig. 2D, 0-s time points). PECs were verified by examining pause escape after adding cold GTP, UTP, and 3′-dATP, the latter of which prevented extension beyond the first A residue. Most of the elongation complexes were active and could be extended to the expected A147, although some complexes arrested one base upstream at U146 (Fig. 2D). This reconstituted PEC faithfully reproduced RNAP pausing at the U144 pause site and pause stimulation by the NGN domain, although the pause duration was shorter than what was observed in promoter-initiated transcription (Fig. 2A), presumably because the pause-stabilizing RNA hairpin is absent from the scaffold. NusA alone did not stimulate pausing of the scaffold-derived PEC but did lead to a synergistic increase in the pause half-life with the NGN domain (Fig. 2D). As this scaffold contained only the native sequence of the U144 pause site as a portion of the ntDNA strand (nt 135–148, Fig. 2B), the finding that this PEC reproduced most of the properties of the native promoter-initiated PEC indicates that the sequence of the ntDNA strand in the paused transcription bubble is sufficient for NusG-stimulated pausing.

Substitutions in positions 137–139 were previously shown to interfere with promoter-initiated NusG-stimulated pausing at the U144 pause site (18). Hence, we tested whether substitution of these three nucleotides in the ntDNA strand would affect pausing. We found that substitution of TTT with CAC resulted in a much shorter pause half-life (oligonucleotide NT2 CAC, Table 1, Fig. 2, B and C). Taken together with our permanganate footprinting results (Fig. 1), we conclude that NusG recognizes this sequence in the ntDNA strand within the paused transcription bubble is sufficient for NusG-stimulated pausing.

**Identification of NusG Residues Required for Pause Stimulation**—The crystal structure of the archaeon *Pyrococcus furiosus* Spt4–Spt5 complex, along with a model of Spt4–Spt5 bound to *P. furiosus* RNAP based on cryo EM single particle reconstruction, was determined (24). The structure of *P. furiosus* Spt4–Spt5 in complex with the RNAP clamp domain was

![FIGURE 2. The NGN domain is sufficient for sequence-specific recognition of the ntDNA strand within the U144 paused transcription bubble.](image-url)

**TABLE 1**

<table>
<thead>
<tr>
<th>DNA and RNA oligonucleotides used for reconstitution of PECs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Oligonucleotide</strong></td>
</tr>
<tr>
<td>tDNA</td>
</tr>
<tr>
<td>T2–11</td>
</tr>
<tr>
<td>T2</td>
</tr>
</tbody>
</table>

| ntDNA | NT2 CAC CCGCCCTCCGTACCATGGTTTTATAGTCTTCCTCTCCATGCC |
| NT2 138BrU | CCGCCCTCCGTACCATGGTTTTATAGTCTTCCTCTCCATGCC |
| NT2 135BrU | CCGCCCTCCGTACCATGGTTTTATAGTCTTCCTCTCCATGCC |

| RNA | RNA 11 UAUAUGUAAAGGCAAGCC |
| RNA 20 | UAUAUGUAAAGGCAAGCC |
| RNA 35 | UAUAUGUAAAGGCAAGCC |

5BrU is in boldface type and underlined.
also determined by x-ray crystallography (25). This study also constructed several models including the bacterial *Thermus thermophilus* RNAP-NusG complex, as well as the eukaryotic yeast Pol II elongation complex with Spt4-Spt5. By using these models, we first built a model of the bacterial RNAP elongation complex with NusG and then replaced *T. thermophilus* NusG with a homology model of *B. subtilis* NusG obtained from the SWISS-MODEL server. We also substituted the ntDNA sequence corresponding to the sequence of the U144 pause site (Fig. 3A). The model shows amino acid residues of NusG adjacent to the ntDNA between positions 135 and 139, allowing sequence-specific contacts with T135, T137, T138, and/or T139 within the U144 PEC (18) (Figs. 1 and 2C). Based on this model, we introduced mutations into several surface-exposed residues of the NGN domain to disrupt sequence-specific interaction with the ntDNA within the U144 PEC. Our strategy was to change specific regions/amino acids of the *B. subtilis* NGN domain with those from *E. coli* NusG because the *E. coli* protein does not respond to the U144 pause signal (Fig. 3B) (15). We tested a total of 21 mutant NGN domains (Fig. 4, A and B, and data not shown). We initially made multiple substitutions in six regions of the *B. subtilis* NGN domain (9–17, 42–60, 77–82, 84–86, 91–99, and 102–118). Mutations in two of these regions led to substantial pausing defects (77–82 and 91–99, Fig. 4B). Because multiple changes in region 77–82 (Y77H/N81S/T82V) completely abolished NusG-stimulated pausing, single amino acid changes within this region were examined. From this analysis, we identified amino acid Thr¹³⁸ as the single most important residue for NusG-stimulated pausing. The adjacent Asn⁸¹ was less critical, whereas the Y77H substitution had little effect on pausing.

Two amino acid regions including residues 9–17 and 84–86 are adjacent to each other in the NGN structure. Multiple substitutions in each of these regions slightly reduced NusG-stimulated pausing, whereas combining the mutations in these two regions severely inhibited pausing. More defined amino acid changes within these two regions revealed His⁹, Thr¹⁰, Gly¹⁸⁴, and Thr¹⁸⁶ as participating in NusG-stimulated pausing. Multiple substitutions in the region 91–99 decreased pausing to the same magnitude as the single T82V substitution. Multiple substitutions in the long β-loop protruding toward the upstream DNA duplex (residues 42–60) had only a marginal effect on pausing. Although the sequence of the C-terminal portion of the NGN domain is the most divergent between *B. subtilis* and *E. coli*, multiple changes in this region (residues 102–118) did not affect NusG-stimulated pausing.

The majority of NGN mutations similarly affected the pause half-life and efficiency. However, substitution of residues 84–86, and especially when substitutions of residues 9–17 were combined with G84R, caused a substantial decrease in the pause efficiency without affecting the pause half-life (Fig. 4B). This observation may suggest that Gly¹⁸⁴ participates in establishment of the pause state but not in pause escape. However, the structural basis for this specificity is not clear (Fig. 3A).

We also tried to convert the *E. coli* NGN domain into a protein that would recognize the U144 pause site by replacing residues 9–17 and 81–99 with those from the *B. subtilis* protein. However, these changes did not result in detectable pause stim-
ulation. Thus, it is apparent that the residues identified in the B. subtilis NGN domain in this study are necessary but not sufficient for pause stimulation.

**NusG-stimulated Pausing**

*FIGURE 4. Structure-function analysis of NGN-stimulated pausing.* A, example of a gel showing the effects of mutations in the NGN domain on RNAP pausing at U144. Promoter-initiated single-round *in vitro* transcription reactions were performed with a trp leader template that was truncated at position 175. Transcription was performed in the presence of 1 μM of WT or mutant NGN domains and 150 μM of each NTP. Reactions were incubated at 23 °C and stopped at the times shown above each lane. Chase reactions (C) were extended for an additional 10 min at 37 °C in the presence of 500 μM of each NTP. Positions of the paused (U144) and run-off (RO) transcripts are shown. The band corresponding to the pause at position 167 was an artifact caused by the close proximity of this base to the end of the DNA template. Pause half-life (t₁⁄₂) and pause efficiency (% (Eff (%)) values for this experiment are shown at the bottom of the gel.

B, effect of B. subtilis NGN domain mutations on pause half-lives and pause efficiencies. Experiments were performed as described in A. WT NGN is in gray. Mutant NGN domains containing substitutions of residues that are responsible for sequence-specific recognition of the ntDNA (green) follow the color scheme in Fig. 3. Values are averages of 2–5 experiments ± S.D.

*FIGURE 5. Crosslinking of the ntDNA strand to full-length NusG or the NGN domain.* The nucleic acid scaffold used in this analysis is shown in Fig. 2B unless indicated otherwise. A, NusG crosslinks to the ntDNA strand at position 138, but not to position 135, when substituted with 5BrU. NusG-independent crosslinks to the β and/or β′ subunits are also shown. B, crosslinking between the tDNA and ntDNA strands in nucleic acid scaffold does not require the presence of RNAP or NusG. Crosslinked products were separated such that free ntDNA, crosslinked tDNA-ntDNA, and crosslinked NusG-ntDNA species were resolved. Species containing ntDNA crosslinked to the RNAP β and β′ subunits did not enter the gel (not shown). C, NusG-ntDNA crosslinking reactions were performed as in B with nucleic acid scaffolds containing 20-mer or 11-mer RNAs. D, the crosslinking efficiency of NGN to ntDNA in which position 138 was replaced with 5BrU is correlated with the pause-stimulating activity of mutant NGN domains. Full-length NusG and the WT NGN domain were used as controls.

NusG Crosslinks to the ntDNA Strand within the Paused Transcription Bubble—Our structural model places the NGN domain in close proximity to the ntDNA in the paused tran-
scriptional bubble (Fig. 3A). Consistent with this model, NusG crosslinked to the ntDNA strand when T138, but not T135, was substituted with 5BrU (NT2 135BrU and NT2 138BrU, Table 1, Fig. 5A). We also observed NusG-independent crosslinks of both modified ntDNA strands to the β and/or β′ subunits of RNAP (Fig. 5A). Because NusG only crosslinked to ntDNA containing 138BrU, all subsequent crosslinking experiments used this modified ntDNA. In addition to the crosslinked species described in Fig. 5A, we observed crosslinks between the ntDNA and tDNA strands. However, these species likely resulted from nucleic acid scaffolds that failed to incorporate RNAP as these species were also observed in the absence of RNAP and NusG (Fig. 5B). We then performed crosslinking experiments with WT and mutant NGN domains. Importantly, the efficiency of crosslinking to mutant NGN domains closely paralleled their pause-stimulating activity (Figs. 4B and 5D).

Spt5 is a eukaryotic homolog of NusG that contains additional C-terminal domains that are absent in bacterial NusG. Spt5 from Drosophila melanogaster requires a nascent transcript longer than 18 nt to stably associate with the Pol II elongation complex. Moreover, this protein contacts the nascent RNA as revealed by protein-RNA crosslinking (26). In contrast, our structural model indicates that the smaller NusG is positioned next to the ntDNA strand but remote from the RNA exit channel (Fig. 3A). To test this aspect of our model, we compared crosslinking of the previously used scaffold containing a 20-mer RNA with a scaffold containing an 11-mer RNA (RNA 11, Table 1). According to all available structural data, RNA 11 must be entirely contained within the RNA-DNA hybrid and the RNAP exit channel. In this scaffold, we used the same ntDNA oligonucleotide (NT2 138BrU) along with oligonucleotide T2–11 as tDNA (Table 1). Although the relative efficiency of interstrand ntDNA-tDNA crosslinking was higher in the PEC with RNA 11, NusG-ntDNA crosslinking was still observed (Fig. 5C). From these data, we conclude that NusG association with the PEC and interaction with the ntDNA strand do not require nascent RNA emerging from the RNA exit channel of the elongation complex.

According to our structural model, a reasonable molecular mechanism of NusG-stimulated pausing is inhibition of RNAP forward movement by simultaneous binding of the NGN domain to RNAP and ntDNA (Fig. 3A). This mechanism has a testable prediction in that WT NGN should bind to the PEC more tightly than mutant NGNs that are unable to interact with the ntDNA in the PEC. To test this prediction, we performed competition experiments between WT and pause-deficient 77–82-mutant NGN domains. In this analysis, PECs were reconstituted from fully complementary tDNA and ntDNA oligonucleotides and a 35-mer RNA that contains the wild-type trp leader sequence, including the U144 pause hairpin. Numbering corresponds to the U144 pause site. PECs formed with the WT NGN domain were challenged with competitor 77–82-mutant NGN at 5-fold molar excess for 0 or 60 s prior to extension. Reactions were stopped at the times shown above each lane (s). Chase reactions (C) were extended for an additional 5 min. Pause half-life values are shown at the bottom of the gel. PECs formed with the 77–82 mutant NGN domain were challenged with competitor WT NGN at 5-fold molar excess for 0 or 60 s prior to extension. Reactions were stopped at the times shown above each lane (s).

FIGURE 6. Wild-type NGN binds more tightly to the PEC than the pause-deficient 77–82 mutant. A, the nucleic acid scaffold used for competition experiments was reconstituted from fully complementary tDNA and ntDNA oligonucleotides and a 35-mer RNA that contains the wild-type trp leader sequence, including the U144 pause hairpin. Numbering corresponds to the U144 pause site. B, PECs formed with the WT NGN domain were challenged with competitor 77–82-mutant NGN at 5-fold molar excess for 0 or 60 s prior to extension. Reactions were stopped at the times shown above each lane (s). Chase reactions (C) were extended for an additional 5 min. Pause half-life values are shown at the bottom of the gel. C, PECs formed with the 77–82 mutant NGN domain were challenged with competitor WT NGN at 5-fold molar excess for 0 or 60 s prior to extension. Reactions were stopped at the times shown above each lane (s). Note that this scaffold formed with low efficiency, which was likely caused by the AU-rich RNA-DNA hybrid (Fig. 6A). Using this scaffold system, we observed exchange between PEC-bound NGN and NGN in solution. Competition of the PEC bound by mutant NGN with 5-fold molar excess of WT NGN and vice versa for 5 min was sufficient to completely displace the originally bound NGN (data not shown). In contrast, challenge by a competitor NGN only during the course of collecting transcription elongation time points (i.e. without preincubation) resulted in little to no displacement of prebound NGN (Fig. 6, B and C, compare pause half-lives of the 0-s preincubation with the corresponding no competitor controls). However, when the PEC was assembled with pause-defective mutant NGN, a 60-s preincubination step with WT NGN competitor prior to pause escape resulted in substantial replacement of the mutant protein as evidenced by the increased pause half-life (Fig. 6C). In the reciprocal experiment in...
which the PEC containing pre-bound WT NGN was challenged with mutant NGN for 60 s prior to pause escape, the pause half-life remained long (Fig. 6B), indicating that the majority of WT NGN was retained in the PEC. Similar results were obtained in promoter-initiated competition experiments (data not shown). Because replacement of PEC-bound NGN with a competitor likely continued during collection of time points of PEC extension, we were unable to calculate accurate dissociation constants. Nevertheless, tighter association of WT NGN is consistent with sequence-specific contacts between B. subtilis NusG and the ntDNA strand within the paused transcription bubble.

Discussion

There is increasing evidence that transcription elongation is regulated by several different effectors that interact with the single-stranded portion of the ntDNA strand that is exposed within the transcription bubble. B. subtilis RNAP pauses at two sites within the trp operon leader region. NusG-stimulated pausing at U107 and U144 participates in transcription attenuation and translation repression mechanisms, respectively, by providing additional time for TRAP to bind to the nascent transcript (13–15). Our studies described here, combined with our previously published results (15, 18), demonstrate that NusG makes sequence-specific contacts with the T-rich ntDNA strand of the U144 paused transcription bubble. We also identified two regions of NusG that are critical for interaction with the T-rich sequence. The fact that this T-rich sequence is conserved in the U107 pause site suggests that NusG stimulates pausing at both sites by essentially identical mechanisms.

E. coli contains the NusG paralog RfaH, which binds specifically to the ops sequence within the ntDNA strand of the transcription bubble (3, 5, 6). The amino acid residues of RfaH that recognize the ops site are located in the same general regions that we identified for NusG.

FIGURE 7. Conservation of pause site recognition by NusG-like elongation factors. Sequence alignment of the B. subtilis NusG and E. coli RfaH NGN domains is shown. The amino acid sequences of two short regions within the NusG NGN domain (shown in red) are primarily responsible for sequence-specific recognition of the U144 trp pause site. The RfaH amino acid residues that recognize the ops pause site (shown in red) are located in the same two regions that we identified for NusG.

FIGURE 8. Model of NusG-stabilized RNAP pausing. During elongation, NusG binds to the β′ clamp helices in close proximity to the ntDNA strand within the transcription bubble, whereas NusA binds to the β flap domain near the RNA exit channel (left). When RNAP reaches the U144 pause site, bases 135, 137, 138 and/or 139 flip out toward the NusG NGN domain. The resulting NusG-ntDNA sequence-specific interaction stabilizes the pause. The pause hairpin further increases the pause duration.

The prolonged pause of the bacteriophage λ late gene operon is another example in which a regulatory factor recognizes the ntDNA strand (28–30). This promoter-proximal pausing event is mediated by σ70-dependent recognition of a −10 promoter-like sequence in the ntDNA strand of the transcription bubble. Pausing at this position plays an important role in the subsequent modification of RNAP by the Q antiterminator protein. Promoter-proximal pausing also occurs during transcription elongation of the E. coli lacZYA operon. As for the situation just described for the λ late gene operon, σ70 recognizes a sequence in the ntDNA strand resembling a −10 promoter element (31). The close proximity of this type of pausing to promoters is a consequence of stochastic dissociation of σ70 such that fewer and fewer elongation complexes contain σ70 as transcription proceeds (32). The pause-stimulating −10-like sequence and the promoter −10 element are probably recognized in a similar fashion (29). σ70 domain 2 forms sequence-specific interactions with the single-stranded −10 DNA element primarily through the −11A and −7T bases, which are flipped out of the single-stranded base stack and buried deep in protein pockets (33).
It was recently shown that the ntDNA strand is a target for co-transcriptional cleavage by type III CRISPR-Cas immunity (where CRISPR indicates "clustered regularly interspaced short palindromic repeat"). In this case, sequence-specific recognition is achieved by ntDNA base pairing with CRISPR RNA. Type III CRISPR-Cas immunity targets only transcribed DNA and is thought to prevent lytic induction of genome-integrated prophages (34). Therefore, the single-stranded portion of ntDNA exposed in the transcription bubble can be recognized by numerous factors in a sequence-specific manner.

B. subtilis NusG contacts the β’ clamp helices of RNAP and the ntDNA strand within the U144 paused transcription bubble. Because transcription elongation requires that RNAP and the DNA template move with respect to one another, simultaneous interaction with both of these components stimulates pausing by preventing forward translocation. We propose that recognition of the B. subtilis pause-stabilizing sequences by NusG involves base flipping of T135, T137, T138, and/or T139 toward the NGN domain (Fig. 8). Capture of these bases by NusG could explain the observed protection of these residues from permanganate oxidation (Fig. 1).

The finding that B. subtilis NusG stimulates pausing at two sites within the same leader transcript suggests that this pausing mechanism may be a common strategy for regulating gene expression. The genome-wide mapping of transcription factors in E. coli revealed that NusG generally associates with elongating RNAP but is absent from promoter-proximal peaks formed by RNAP. Moreover, the amount of NusG gradually rises as transcription progresses to a level that exceeds other factors on most transcriptional units (32). Therefore, NusG has the capacity to stimulate RNAP pausing at any position of any transcription unit if its recognition sequence is present in the ntDNA. Although E. coli NusG is known as an antipauasing factor, future studies may reveal specific pause-stabilizing recognition sequences for this protein. In contrast to NusG, its eukaryotic homolog Spt5 requires a nascent transcript to stably associate with the Pol II elongation complex. However, this interaction is unlikely to preclude additional Spt5 interactions with the ntDNA. Instead, the more complex eukaryotic transcription machinery may require Spt5-dependent regulation in response to specific signals in both ntDNA and nascent RNA.

Author Contributions—A. V. Y. and P. B. conceived the project. A. V. Y. performed all experiments. K. S. M. generated the structural model of NusG bound to the elongation complex. A. V. Y. and P. B. wrote the manuscript. P. B. supervised the project and secured funding. All authors analyzed the results and approved the final version of the manuscript.

References
NusG-stimulated Pausing


