GreA and GreB proteins revive backtracked RNA polymerase \textit{in vivo} by promoting transcript trimming

Francine Toulmé, Christine Mosrin-Huaman, Jason Sparkowski\textsuperscript{1}, Asis Das\textsuperscript{1}, Marc Leng and A.Rachid Rahmouni\textsuperscript{2}

Centre de Biophysique Moléculaire, CNRS, rue Charles Sadron, 45071 Orléans cédex 2, France and \textsuperscript{1}Department of Microbiology, University of Connecticut Health Center, Farmington, CT 06030, USA

\textsuperscript{2}Corresponding author
e-mail: rahmouni@cnrs-orleans.fr

The GreA and GreB proteins of \textit{Escherichia coli} show a multitude of effects on transcription elongation \textit{in vitro}, yet their physiological functions are poorly understood. Here, we investigated whether and how these factors influence lateral oscillations of RNA polymerase (RNAS) \textit{in vivo}, observed at a protein readblock. When RNAS is stalled within an (ATC/TAG)\textsubscript{3} sequence, it appears to oscillate between an upstream and a downstream position on the template, 3 bp apart, with concomitant trimming of the transcript 3' terminus and its re-synthesis. Using a set of mutant \textit{E.coli} strains, we show that the presence of GreA or GreB in the cell is essential to induce this trimming. We show further that in contrast to a ternary complex that is stabilized at the downstream position, the oscillating complex relies heavily on the GreA/GreB-induced ‘cleavage-and-restart’ process to become catalytically competent. Clearly, by promoting transcript shortening and re-alignment of the catalytic register, the Gre factors function \textit{in vivo} to rescue RNAS from being arrested at template positions where the lateral stability of the ternary complex is impaired.

\textit{Keywords:} elongation factors/GreA/GreB/RNA polymerase/transcription elongation complex

Introduction

RNA polymerases (RNAPs) carry out transcript elongation in a discontinuous manner. Frequently, DNA sequences through which the polymerase must pass to complete an RNA chain lead the enzyme to stop RNA synthesis for varying lengths of time before resuming elongation. This process of RNAP pausing is physiologically significant for both positive and negative control of transcription elongation (Das, 1993; Henkin, 1996; Landick \textit{et al.}, 1996; Uptain \textit{et al.}, 1997; Artsimovich and Landick, 2000). In some instances \textit{in vitro}, a fraction of the paused polymerases eventually transforms into an arrested state in which the ternary complex neither elongates nor dissociates (Davenport \textit{et al.}, 2000 and references therein). According to current models, this loss of catalytic activity results from a backward translocation of RNAP along the DNA and RNA chains (Landick, 1997; von Hippel, 1998; Nudler, 1999). The upstream movement shifts the transcription bubble, the RNA–DNA hybrid and the catalytic center away from the 3' end of the transcript, which becomes extruded out of the polymerase (Reeder and Hawley, 1996; Komissarova and Kashlev, 1997a,b; Nudler \textit{et al.}, 1997; Samkurashvili and Luse, 1998; Sidorenkov \textit{et al.}, 1998). Thus, unless the sliding process is reversed, resumption of RNA synthesis requires trimming of the transcript to generate a new 3' hydroxyl group in register with the catalytic center of the enzyme.

A special class of elongation factors, which includes GreA and GreB for the bacterial RNAP and SII for the eukaryotic RNAP II, stimulates transcript cleavage within ternary complexes that are artificially halted by NTP starvation (Borukhov \textit{et al.}, 1992, 1993; Reines, 1992, 1994; Reines \textit{et al.}, 1992; Kassavetis and Geiduschek, 1993). These \textit{in vitro} observations raised the attractive possibility that the Gre/SII family of elongation factors may play a key role \textit{in vivo} (i) to mitigate pausing and suppress elongation arrest and (ii) to enhance transcription fidelity (Erie \textit{et al.}, 1993; Jeon and Agarwal, 1996; Thomas \textit{et al.}, 1998). To date, however, evidence that GreA and GreB as well as SII act as cleavage factors inside the cells is still lacking. Ternary transcription complexes made with a variety of purified RNAPs have been shown to possess substantial endonucleolytic cleavage activity in the absence of the known stimulatory factors (Surratt \textit{et al.}, 1991; Hagler and Schuman, 1992; Wang and Hawley, 1993; Rudd \textit{et al.}, 1994; Tschochner, 1996; Sastry and Ross, 1997; Chedin \textit{et al.}, 1998). In particular, experiments with \textit{Escherichia coli} RNAP purified from a greA\textsuperscript{-}greB\textsuperscript{-} double mutant indicated that the cleavage reaction is inherent to the polymerase and that Gre factors simply enhance this intrinsic property (Orlov \textit{et al.}, 1995). Nevertheless, the biological function of the transcript trimming activity of any RNAP has yet to be revealed.

We have recently shown that a transcription elongation complex, halted within an (ATC/TAG)\textsubscript{n} sequence due to a readblock imposed by the lac repressor in \textit{E.coli}, oscillates between an upstream and a downstream position on the template, with accompanying transcript cleavage and re-synthesis (Toulmé \textit{et al.}, 1999). We have now investigated the influence of GreA and GreB on RNAP oscillations in this system. Results reported below provide direct evidence for the transcript cleavage reaction mediated by GreA and GreB inside the cell. They also show that the Gre factor-induced transcript shortening is essential for RNAP to read through template sequences that impair the lateral stability of the ternary complex.
ternary complex halted within an (ATC/TAG)_n sequence, which potentially generates an unstable RNA–DNA hybrid, is distributed between a downstream (–6) and an upstream (–9) translocated position (Figure 1, plasmid pATC6a). In contrast, a template sequence that should yield a stable RNA–DNA hybrid holds the ternary complex fixed in the downstream location (plasmid pATC6b).

We envisaged that the multiple forms of the ternary complex observed within the (ATC/TAG)_n stretch are in a dynamic equilibrium, where the readblocked polymerase oscillates between two positions on the template with accompanying transcript cleavage and re-synthesis. According to this scenario, the polymerase first elongates the transcript to the point where the operator-bound lac repressor becomes a physical barrier. Within such a readblocked complex, the catalytic center of the enzyme is in register with position –6 (the location of the 3′ terminal RNA nucleotide) with respect to the upstream edge of the operator motif. When the RNA–DNA hybrid within the complex is relatively strong (5 rC–dG bp in the putative 8 bp heteroduplex), the polymerase remains fixed at this single site on the template. However, when the complex contains a weak hybrid (3 rC–dG bp in the heteroduplex), the polymerase becomes prone to slide backward. The backsliding of RNAP relocates the catalytic center in register with position –9, where transcript cleavage occurs. This generates the upstream translocated complex. Re-elongation of the transcript from the new 3′ end brings RNAP back to the downstream, readblocked location.

Conceivably, transcript cleavage within the backslided ternary complex is an intrinsic activity of polymerase, or

Fig. 1. Schematic models of a stable and an oscillating elongation complex halted in vivo by the lac repressor bound to its operator. The middle section of the figure shows the relevant part of plasmids pATC6a and pATC6b with the non-template strand sequence of the repeats. These plasmids, which have been described previously (Toulmé et al., 1999), are pKK232-8 derivatives that carry the β-lactamase (Amp) and chloramphenicol acetyltransferase (CAT) genes. The transcription of the cat gene initiated at the constitutive hisR promoter goes through the repeats and the operator sequences that are inserted at position +40 relative to the transcription start site. Positions –6 and –9 relative to the upstream edge of the operator motif are also indicated, as the mRNA start site (right-angled arrow). See text for details.

Results

GreA and GreB induce transcript cleavage within an oscillating ternary complex in vivo

To investigate the cellular function of the Gre factors, we analyzed in vivo the structural behavior of an RNAP elongation complex readblocked on a reporter plasmid (Figure 1). On this template, transcription is initiated from a constitutive promoter and the elongation complex is stalled by a physical readblock, imposed by the lac repressor bound to its operator. By a combination of in situ DNA footprinting and RNA 3′ end mapping experiments, we have shown previously (Toulmé et al., 1999) that the
the RNA trimming is induced by Gre factors. To obtain evidence to support these possibilities, we extended the footprinting analyses with E. coli cells that lack functional GreA, GreB or both. The pATC6a plasmid that harbors the oscillation-inducing template sequence was moved into a set of four isogenic strains: wild type, greA::KanR, ΔgreB::CamR and the greA− greB− double mutant. We first mapped the 3′ ends of the truncated transcripts generated by the readblocked ternary complex.

Cellular RNAs were extracted from the transformed cells and the RNA 3′ ends were determined by S1 nuclease protection experiments (see Materials and methods). In agreement with previous results (Toulmé et al., 1999), the RNAs extracted from wild-type cells grown in the absence of the inducer isopropyl-β-D-thiogalactopyranoside (IPTG) contained short transcripts with 3′ ends distributed primarily between positions −6 and −9 with respect to the lac operator (Figure 2A). This signature of the apparent ‘cleavage-and-restart’ process within the oscillating complex was also observed for the RNAs extracted from either the greA− or the greB− mutants. In striking contrast, a major band corresponding to the 3′ end at position −6 was detected for the RNAs produced in the double-mutant strain. The intensity of this band is approximately equal to the sum of the two signals (positions −6 and −9) observed for either the wild type or single mutants (Figure 2B). These results are consistent with the hypothesis that the RNAP readblocked within the (ATC/TAG)ₙ tract backslides, and that it resumes elongation upon transcript shortening promoted by GreA and GreB.

The apparent transcript cleavage activity in the readblocked ternary complex is more efficient with GreA than with GreB. A quantitative analysis of the S1 nuclease protection patterns obtained with the single mutants revealed a reproducible, albeit modest, difference between the distributions of the RNA 3′ ends (see densitometric scans in Figure 2B). Whereas the 3′ ends in the greB− mutant are divided in nearly equal amounts between positions −6 and −9, as was observed with the wild-type strain, the RNAs produced in the greA− mutant have more 3′ ends at position −6.

Another diagnostic for the cleavage and re-synthesis process within the oscillating complex is obtained by analyzing the methylation status of the G residue at position −6 on the template strand. The in situ dimethylsulfate (DMS) probing data, shown in Figure 3, reveal the readblock-dependent hypermethylation of the G-6 residue in wild-type cells and single mutants, but clearly not in the double mutant. The extent of methylation of the G-6 residue in the double mutant is only slightly greater than that of the surrounding G residues. These results strongly support our previous suggestion that the hyper-reactivity of the G-6 residue with DMS is associated with the dynamic equilibrium, the template base residue being repeatedly blocked in a highly exposed configuration during the re-incorporation of the rC residue into the transcript (Toulmé et al., 1999). Thus, the absence of both GreA and GreB renders the stalled elongation complex transcriptionally quiescent by impairing the dynamic process of cleavage and re-synthesis between the −6 and −9 positions.

Interestingly, these in situ DMS probing experiments (Figure 3) also show a slightly lower hypermethylation of the G-6 residue in the greA− cells compared with the greB− or wild-type strains (densitometric scans not shown). In agreement with the S1 mapping data described above, these results emphasize a lower frequency of rC-6 re-incorporation in the greA− strain due to the reduced efficiency of transcript cleavage.

**The catalytic competence of the backslid RNA polymerase requires functional GreA or GreB**

The molecular properties of the ternary complex generated in the greA− greB− double mutant with plasmid pATC6a could be readily explained if the readblocked RNAP is no longer sliding back and forth within the (ATC/TAG)ₙ tract, but instead is trapped in the downstream position.
Escherichia coli cells bearing pATC6α were treated with CAA, and following plasmid DNA extraction the modified bases on the non-template strand were revealed by primer extension. As shown in Figure 4, the CAA reactivity patterns of pATC6α modified in the wild-type and double-mutant strains are virtually identical, although the modifications at the upstream and downstream margins of the footprint differ slightly between the two strains. The CAA reactivities are spread over a region of ~14 bases located upstream from position ~6. This apparent accessibility to the probe of a large region of the non-template strand must result from the superposition of the different forms of the complex that are in equilibrium. As shown above (Figure 2), the ternary complexes of pATC6α in the greA− greB− double mutant have a single 3’-ended transcript (position ~6). Thus, these CAA probing results indicate that in the absence of functional Gre factors, the ternary complexes readblocked within the (ATC/TAG)ₙ stretch do slide back and forth between the upstream and downstream locations; however, they do not shorten the 3’ terminal part of the transcript. Notably, the slight shift of the CAA footprint towards the upstream positions observed in the double mutant might reflect the fact that the oscillating ternary complexes spend more time at the backward location when GreA and GreB are absent (Figure 4). In the view where the backward sliding of the polymerase extrudes the RNA 3’ tip from the active center, the upstream form of the complex should be catalytically inactive. Evidence supporting this notion is presented below.

The operator-bound lac repressor is expected to cause only a transient block to the emergent elongation complex. Once the two complexes are in close apposition on the DNA, the duration of the block is a function of the dwelling time of the lac repressor on its binding site. However, the efficiency with which the polymerase will read through the operator motif, following repressor dissociation, depends upon whether the ternary complex is able to escape rapidly from this position before the lac repressor re-binds DNA. Thus, the level of transcription readthrough under repressing conditions should be related to the catalytic competence of the halted ternary complex.

Therefore, we measured the amounts of downstream cat mRNA produced in the different strains harboring pATC6α. A 32P end-labeled oligonucleotide complementary to the early part of the cat message was used to quantitate mRNA by primer extension with reverse transcriptase. A second oligonucleotide was also included to detect the plasmid-encoded β-lactamase transcript (bla) for normalization. Under repressing conditions, a fraction of RNAPs did transcribe beyond the operator motif, and this level of readthrough was clearly affected by the absence of both GreA and GreB (Figure 5A). Curiously, the gre mutations showed a defect in cat gene expression that is independent from the lac repressor impediment, since a reduction in cat mRNA was also observed under inducing conditions. We have not addressed the molecular basis of this defect. We note, however, that this could be due to a reduced frequency of RNA chain initiation in the gre mutants (Hsu et al., 1995; A.Das et al., manuscript in preparation). Alternatively, in the absence of GreA and GreB, RNA chain elongation could have been delayed at some natural pause sites within the early part of the cat

Indeed, similar characteristics were found previously in wild-type cells for a ternary complex that is stably positioned at the downstream location (as in the case of the pATC6b variant). Alternatively, we imagine that in the absence of GreA and GreB, the ternary complex in pATC6a is still repeatedly switching between the upstream and downstream positions without breaking the 3’ tip of the transcript. A distinguishing feature between the oscillating and non-oscillating elongation complexes is the apparent size of the transcription bubble. Whereas in the laterally stabilized complex the DNA bubble is limited to 8 bp, the oscillating complex is characterized by an appropriately larger DNA bubble (Toulmé et al., 1999). Therefore, we extended our comparative studies by analyzing the accessibility of the non-template strand to the single-strand-specific chemical probe chloroacetaldehyde (CAA).
gene. To evaluate specifically the readblock-dependent effects of the Gre factors, we normalized, for each strain, the level of cat mRNA produced under repressing conditions to that obtained under inducing conditions. This analysis revealed a clear defect of RNAP in overcoming the readblock in the absence of GreA and GreB (Figure 5B). Whereas the efficiency of readthrough was almost unaffected in the greB mutant compared with the wild-type cells, the greA strain displayed a 30% reduction of cat mRNA. By comparison, the readthrough was reduced by 60% in the double mutant.

We infer that the ternary complex in pATC6a slides back and forth between a catalytically dormant (upstream) and an active form (downstream). Upon repressor dissociation, only the active form escapes before the lac repressor re-binds. By promoting transcript trimming and realignment of the catalytic register, the Gre factors reactivate the backtracked ternary complex more efficiently than do the oscillations without cleavage. Our conclusion is clearly supported by the results with the pATC6b variant (Figure 5B). In this case, where the ternary complex is fixed in the downstream location (active form), the efficiency of transcription readthrough is virtually unaffected by the absence of the two Gre factors.

Discussion

Transcription processivity, essential for gene expression and control, relies on the perplexing grips of RNAP that hold the template DNA and the nascent transcript in a stable but flexible manner. The stability of these grips prevents the random dissociation of the ternary complex, while their flexibility enables the catalytic center to rapidly reposition itself (15–50 ms) at every cycle of nucleotide addition. During typical elongation, the polymerase has little or no reason to march backward. Yet, RNAP does slide backward at certain template sites, when it is deprived of nucleotides, or when RNAP encounters some natural pause sites or a physical barrier imposed by a protein bound on the DNA. The backsliding event renders RNAP catalytically incompetent owing to the out-of-register catalytic center. As a result, transcription stalls for varying periods of time until RNAP regains catalytic activity. The experiments presented in this paper show that elongation factors GreA and GreB play a significant role in the cell to revive RNAP from its backslid, catalytically incompetent state. They do so by promoting transcript cleavage, which generates a new 3' terminus of the transcript in register with the catalytic center (Figure 1).

We have shown that the backslid RNAP fails to trim the RNA 3' terminus in vivo when both GreA and GreB are absent. This contrasts with previous in vitro results demonstrating that both E.coli RNAP and eukaryotic RNAP II possess intrinsic transcript cleavage activity (Rudd et al., 1994; Orlova et al., 1995). Notably, in these in vitro experiments, transcript hydrolysis at internal positions was induced by specific conditions, such as NTP starvation and high pH or the presence of pyrophosphate activity. These studies suggested that the endonucleolytic cleavage reaction is probably catalyzed by the same active center that performs RNA synthesis. Our findings do not exclude this possibility, but they clearly indicate that within the cellular environment of E.coli, the interaction of GreA or GreB with the ternary complex is essential for the cleavage reaction to occur. We suspect that this rule is likely to hold true for eukaryotes as well.

GreA and GreB are closely related proteins that share a substantial amino acid sequence homology, have a similar structural organization and are conserved in the bacterial kingdom (Koulich et al., 1997). However, the two proteins appear to act differently within ternary complexes in vitro. Whereas GreA-induced transcript cleavage produces short RNAs, two or three nucleotides long, GreB (like SII) produces 2–18 nucleotide RNA fragments (Borukhov et al., 1992, 1993; Reines, 1992; Reines et al., 1992). Indeed, our in vivo results presented here reveal a differential sensitivity of RNAP for the Gre factors, and they provide evidence that this differential specificity is biologically significant. While the two factors clearly substitute for each other in reviving backtracked RNAP in our system, GreA is more efficient than GreB. Apparently, GreA is more specialized for modulating RNAP where backtracking is limited over a short distance. The molecular mechanism underlying this specialization is still unclear. Recent experiments suggest that the functional property of the two Gre proteins is determined in part by the size of the respective basic patch residing on the N-terminal coiled-coil domain (Kulish et al., 2000).

The ‘Zipping’ model of transcript elongation (Reeder and Hawley, 1996; Komissarova and Kashlev, 1997a,b; Nuider et al., 1997; Toulmé et al., 1999) holds that the lateral stability of the ternary complex at each template position is governed by the direct competition between the DNA–DNA and RNA–DNA base pairs that can form at the upstream and downstream branching points of the transcription bubble. Accordingly, we expect that the back-and-forth oscillations of the ternary complex over short distances, such as that operational in our experimental system, are relatively common under the dynamic conditions of transcription in vivo. RNAP oscillations should be induced at template sequences where the competition between the base pairs at the two branching points of the bubble is unfavorable to elongation. Furthermore, misincorporation, which hinders further catalysis, should favor a backward slide of RNAP. Our results show clearly that the oscillating complex relies heavily on the ‘cleavage-and-restart’ mechanism mediated by Gre factors to escape from the protein readblock. In the absence of GreA and GreB, the ternary complex spends time switching between active and inactive isoforms. Thus, the Gre proteins are processivity factors that preclude RNAP from being retarded at positions where the lateral stability of the ternary complex is impaired. They probably function to prevent RNAP from extensive backtracking that would lead to an irreversible arrest. Finally, the Gre factors must also play a significant role in editing the misincorporated RNAs in the cell. It would be surprising if bacteria do not utilize GreA and GreB to control specific genes by suppression of transcription arrest and RNA editing.

The disruption of the two gre genes in E.coli is not lethal under normal growth conditions, but renders the cells thermosensitive (Orlova et al., 1995). It is tempting to speculate that the lateral oscillations of the ternary complex are aggravated by high temperature, which in turn makes the cell more dependent on the Gre factors for
proper gene expression and control (A. Das et al., manuscript in preparation). We imagine that the high temperature defect might be amplified by some genetic alterations in RNAP that favor its propensity for backward sliding, either directly or indirectly. This would certainly explain why certain RNAP mutants confer high temperature lethality that is rescued by an elevated level of GreA (Spurkowski and Das, 1990, 1992).

Materials and methods

Enzymes, chemicals and oligonucleotides

Restriction enzymes and T4 polynucleotide kinase were obtained from New England Biolabs. The Kleenow fragment of DNA polymerase I was purchased from Amersham. Superscript RNase H—Reverse Transcriptase from M-MLV, S1 nuclease and T4 DNA ligase were from Life Technologies. DMS and most chemicals, including antibiotics, were from Sigma. CAA was bought from Fluka Chemie and double distilled before use (boiling point 78–80°C). Unlabeled dNTPs were from Boehringer Mannheim, whereas [γ-32P]dATP, [α-32P]dATP and all the oligonucleotides were from Amersham.

Bacterial strains and plasmids

The four Es.coli recipient strains used are isogenic to W3110 (E.coli Genetic Stock Center). The AD8782, AD8786 and AD8775 strains harbor the greA (greA: KanR), the greB (AgrB2:CamR) and the greA: greB double mutations, respectively (A. Das et al., manuscript in preparation). In all subsequent in vivo experiments, the strains were grown with the appropriate antibiotics at the following concentrations: kanamycin, 18 μg/ml; chloramphenicol, 15 μg/ml.

The pATC6a and pATC6b plasmids described previously (Toulmé et al., 1999) are pKK232-8-based plasmids. These constructs contain a constitutive promoter that drives transcription of the cat gene through the (ATC/TAIG), and the lac and the luc operator motif (Figure 1). The introduction of either of these plasmids into the strains described above is via selection on agar plates containing 100 μg/ml ampicillin in addition to the aforementioned antibiotics. A second plasmid (pCA184Q) that overproduces lac repressor was then introduced in these transformants by selecting tetracycline (16 μg/ml) and ampicillin resistance. The pCA184Q plasmid was created by subcloning the lacI gene from pCA177IQ (Toulmé et al., 1999) into the PvuII (position 515) and SstI (position 2863) restriction sites of pACYC184 (New England Biolabs), thus deleting the cat gene. In general, all DNA manipulations and transformations into E.coli cells were performed according to standard procedures (Sambrook et al., 1989).

In situ DNA footprinting and RNA analyses

The in situ DNA probing with DMS or CAA and the subsequent analyses of the modifications by hot piperidine cleavage (DMS) or primer extension with the Kleenow fragment of DNA polymerase I (CAA) were carried out as described (Guérin et al., 1996; Toulmé et al., 1999). The RNA extraction and the 3′ end mapping with S1 nuclease protection experiments were also performed exactly as reported previously (Toulmé et al., 1999).

Quantitative analyses of the cat and bla transcripts were carried out simultaneously with two 3P end-labeled primers that anneal between positions 4850 and 4868 for the bla gene and between positions 248 and 265 for the cat gene (the positions are those of the original vector pKK232-8). Twenty micrograms of total RNAs were mixed in water with 0.1 pmol of each labeled primer in a volume of 10 μl. The samples were heat denatured at 90°C for 2 min and chilled on ice. Four microliters of 5× reverse transcription buffer (250 mM Tris–HCl pH 8.3, 375 mM KCl, 15 mM MgCl2), 2 μl of 100 mM dithiothreitol and 5 U of RNase inhibitor (RNAGuard™ from Amersham) were then added prior to hybridization at 45°C for 20 min. cDNA synthesis was initiated by the addition of a mixture containing 4 μl of dNTPs (5 mM each) and 100 U of reverse transcriptase from M-MLV, and incubation was continued at 45°C for 1 h. Nucleic acids were then ethanol precipitated, dried, and finally resolved on a 6% denaturing polyacrylamide gel. Note that the cat primer also hybridizes to the chromosome-encoded cat transcript produced in the AD8786 and AD8775 strains. However, the resulting cDNAs are easily identified by their smaller size and are not present on the part of the gel shown in Figure 5A.

Acknowledgements

This article is dedicated to the memory of Marc Leng who died on May 7, 2000. We are grateful to Marc Boulvillain and Martine Guérin for helpful discussions. This work was supported in part by the Association de la Recherche sur le Cancer, la Ligue Contre le Cancer (comité du Loiret) and by the USPHS Grant (GM28946) from the National Institutes of Health to (A.D.).

References


Received September 14, 2000; revised and accepted October 24, 2000.