Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 8899–8902, October 1992
Biochemistry

GreA protein: A transcription elongation factor from *Escherichia coli*

(ternary complex/transcript cleavage/elongation arrest/stalling sites)

SERGEI BORUKHOV*, ANDREY POLYAKOV*†, VADIM NIKIFOROV*†, AND ALEX GOLDFARB*‡

Department of Microbiology, Columbia University College of Physicians and Surgeons, New York, NY 10032

Communicated by Michael J. Chamberlin, June 30, 1992 (received for review May 15, 1992)

ABSTRACT **A protein identified as the 158-amino acid product of the *greA* gene was isolated from *Escherichia coli*. When added to a halted ternary transcription complex, the GreA protein induced cleavage and removal of the 3′ proximal dinucleotide from the nascent RNA. The new 3′ terminus generated by the cleavage could be extended into longer transcripts. GreA-mediated cleavage of a transcript appears to permit a ternary complex to resume transcription from a state of indefinite elongation arrest induced by a specific DNA site. The GreA protein tended to interact with RNA polymerase during purification and recycled between RNA polymerase molecules in the course of the *in vitro* cleavage reaction. Similar biochemical activities have been reported in eukaryotic RNA polymerases, indicating that transcript cleavage and restart of elongation may be a general transcriptional mechanism.**

RNA polymerase is the key enzyme of gene expression and serves as the ultimate target for a myriad of regulatory mechanisms. The search for factors that modify and regulate basic biochemical reactions of RNA polymerase is central to the study of gene regulation. The *greA* gene of *Escherichia coli* has been implicated in some vital aspect of transcription by virtue of its ability to suppress, at high copy number, a temperature-sensitive mutation in the RNA polymerase β subunit (1, 2). It was proposed that the *greA* product, which on the basis of DNA sequence should consist of 158 amino acids, directly interacts with the RNA polymerase molecule. We report here a serendipitous finding that GreA is in fact responsible for the recently discovered biochemical reaction of nascent transcript cleavage.

The transcript cleavage reaction first described for *E. coli* RNA polymerase by Chamberlin and coworkers (3) leads to cleavage and removal of a 3′ terminal oligonucleotide from the RNA component of a halted elongation complex. After such cleavage, elongation can resume from the newly generated 3′ transcript end if NTPs are present. Similar reactions have since been found in eukaryotic systems and depend on the elongation factor SII (4, 5). In *E. coli*, the cleavage activity was proposed to be an intrinsic property of RNA polymerase. We demonstrate here that the cleavage reaction is in fact induced by a transacting factor present as a contaminant in standard preparations of *E. coli* RNA polymerase. The factor has been purified to apparent homogeneity from cell extracts and identified as the product of the *greA* gene.

MATERIALS AND METHODS

RNA Polymerase. "Standard preparation" RNA polymerase was purified from *E. coli* MRE600 (Grain Processing, Muscatine, IA) according to ref. 6 with minor modifications. "Recombinant" RNA polymerase was reconstituted from individual subunits (7) by using slightly modified renaturation conditions described in ref. 8.

***rrnB P1* Ternary Complex.** The stable ternary complex containing hexamer transcript CpA**pC**pCpA**pC** (here and elsewhere in the paper boldface letters indicate radioactive phosphate) was prepared by using the isolated 202-base-pair DNA fragment (endpoints of −152 and +50) carrying the ribosomal *rrnB P1* promoter (9). One and six-tenths micrograms (13 pmol) of the DNA fragment was incubated with 35 μg (87 pmol) of RNA polymerase in a 30-μl reaction mixture containing 40 mM Tris acetate (pH 7.9), 30 mM KCl, 10 mM $MgCl_2$, bovine serum albumin (1 mg/ml), 1 mM CpA, 1 μM [α-^{32}P]CTP (3000 Ci/mmol; 1 Ci = 37 GBq), and 10 μM ATP for 10 min at 37°C. The complex was purified on G-50 Quick Spin columns (Boehringer) in the same buffer without NTP.

RNA Chain Extension and Cleavage Reactions. The purified CpA**pC**pCpA**pC** ternary complex was the starting material in all experiments. The reactions were performed in the spin column buffer (the reaction buffer described above but without NTP) in 10-μl samples containing ≈0.3 pmol of the ternary complex. The incubation at 37°C was for 10 min unless indicated otherwise. Where indicated, combinations of unlabeled ribonucleoside triphosphates were added to a final concentration of 10 μM each, as well as the specified amounts of GreA-containing fractions. After incubation, 10 μl of formamide buffer was added, and the samples were analyzed in 23% polyacrylamide/urea gels as described (10).

Purification of GreA. All procedures were at +4°C. The assay for GreA activity in the fractions was the conversion of the heptamer CpA**pC**pCpA**pC**pU into the pentamer CpA-**pC**pCpA by using the ternary heptamer complex as substrate, as is illustrated for the lysate fraction in Fig. 1*C* (see *Results*). Thirty grams of *E. coli* MRE600 cells was suspended in 120 ml of lysis buffer (0.1 M Tris·HCl, pH 7.9/0.1 M NaCl/10 mM EDTA) containing 0.1 mM phenylmethylsulfonyl fluoride and 0.5% (wt/vol) sodium deoxycholate and disrupted by repeated sonication. After removal of debris by centrifugation (15,000 × *g* for 40 min; the lysate fraction), 2 vol of ethanol was added; the precipitate formed at −20°C was collected by centrifugation, washed with 70% ethanol, and extracted with 120 ml of the lysis buffer. NaCl was added to the extract (final concentration of 1 M) followed by the addition of Polymin P [to 0.3% (vol/vol)], and the precipitate formed was removed by centrifugation and discarded. $(NH_4)_2SO_4$ (final concentration of 1.5 M) was added to the Polymin supernatant; the solution was cleared by centrifugation and applied to a 1.6 × 20 cm phenyl-Toyopearl TSK 650M (Supelco) column equilibrated with TE (0.05 M Tris·HCl, pH 7.9/10 mM EDTA) containing 1.5 M $(NH_4)_2SO_4$. After washing the column with 120 ml of the same buffer, the material was eluted by 400 ml of a 1.5 M to

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Public Health Research Institute, 455 First Avenue, New York, NY 10016.
†On leave from the Institute of Molecular Genetics, the Russian Academy of Sciences, Moscow, Russia.
‡To whom reprint requests should be addressed.

0.0 M $(NH_4)_2SO_4$ gradient in TE. The fractions containing active material were pooled and concentrated by ultrafiltration on an Amicon Centriprep-30 to a final volume of ≈1.0 ml. The material (two 500-μl samples) was fractionated on an FPLC Superose 12 10/30HR column in TE containing 0.5 M NaCl. The fractions containing active material were pooled, concentrated as above, diluted 5-fold with TE, and applied on an FPLC Mono Q 5/5HR column; the material was eluted by a 30-ml 0–0.6 M NaCl gradient in TE. The peak containing active material (eluting at ≈0.3 M NaCl) was collected, and the material was precipitated with 2 vol of ethanol as above and dissolved in 8 M urea in TE. The material was then applied to a Mono Q column in TE containing 8 M urea and eluted with 30 ml of a 0–0.4 M NaCl gradient in TE containing 8 M urea. Peak fractions containing active material (eluting at ≈0.12 M NaCl) were pooled, and the urea was removed by ultrafiltration in TE in a Centricon-30 device (Amicon). The material (total yield ≈ 100 μg) was stored at a concentration of 1 mg/ml in TE containing 40% glycerol at −20°C.

Protein Analysis. The electrophoresis in Tris/Tricine buffer system was according to ref. 11. N-terminal sequencing of the M_r 19,000 protein band eluted from the gel was performed as described (12). Amino acid composition analysis of the final M_r 19,000 preparation was carried out at the Columbia University Core Protein Facility. Protein concentration was determined using a Bio-Rad protein assay kit with bovine serum albumin as standard. GreA molarity was calculated from the published amino acid sequence (2).

RESULTS

RNA Cleavage and Restart in the Early Ternary Complex on the *rrnB P1* Promoter. This investigation was prompted by the observation of Surratt *et al.* (3) that the heptameric transcript present in a ternary promoter complex undergoes spontaneous internal cleavage followed by the loss of the 3′ proximal dinucleotide and restart of elongation from the newly generated 3′ terminus of the pentameric cleavage product. These authors used an artificially engineered variant of the *E. coli tac* promoter, which, in contrast to most of *E. coli* promoters, formed unusually stable early ternary complexes, making these amenable to structural analysis. We took advantage of the recent observation that the ribosomal RNA promoter *rrnB P1* also leads to an unusually stable early ternary complex (ref. 9; S.B., V. Sagitov, and A.G., unpublished data), thus enabling us to test the conclusions of Surratt *et al.* (3) in the context of a natural promoter. The initial transcribed sequence of *rrnB P1* is

CACCACUG →.

Our experimental system included the standard preparation RNA polymerase purified from *E. coli* by conventional techniques (see *Materials and Methods*) and the isolated *rrnB P1* promoter fragment. In the presence of the dinucleotide primer CpA, ATP, and [^{32}P]CTP, this template directed the formation of the RNA hexamer CpA**p**CpCpA**p**C, which remained in the ternary complex after size-exclusion purification (Fig. 1*A*, lane 4). The addition of UTP to such a complex led to the transient formation of the heptamer CpA**p**CpCpA**p**CpU, which was rapidly cleaved into the pentamer CpA**p**CpCpA and the dinucleotide **p**CpU (lane 5). The pentamer remained in the ternary complex since it could be converted back into the heptamer in the presence of CTP and UTP (lane 6) and into the octamer in the presence of CTP, UTP, and GTP (lane 7). The **p**CpU dimer, however, was released and could be removed by size-exclusion chromatography (lane 7). It should be noted that in the presence of NTP (lanes 6 and 7), the chain extension reaction completely masked the cleavage reaction, suggesting that in normal transcription the latter would be kinetically insignificant unless the complex is stalled. The cleavage reaction was Mg^{2+} dependent (data not shown). Thus, the RNA cleavage-and-restart reaction on the *rrnB P1* promoter had exactly the same features as the one originally described by Surratt *et al.* (3).

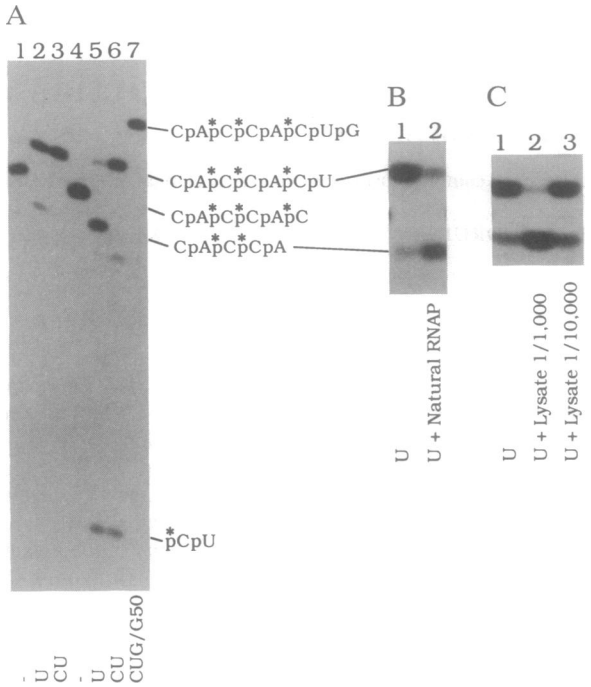

FIG. 1. Cleavage of the heptameric transcript in the ternary complex by a transacting factor. The gel autoradiogram shows RNA oligomers whose identity was established by using differential ^{32}P-labeling, chain-terminating substrates, and dinucleotide standards (S.B., V. Sagitov, and A.G., unpublished data). Asterisks indicate radioactive phosphates. In all experiments, which were performed as described in *Materials and Methods*, G-50-purified ternary complex carrying the hexamer CpA**p**CpCpA**p**C was used as the starting material. (*A*) The hexamer-carrying complex made with recombinant (lanes 1–3) or standard preparation (lanes 4–7) RNA polymerase was analyzed directly (lanes 1 and 4) or after additional incubation with nonradioactive UTP (lanes 2 and 5), UTP plus CTP (lanes 3 and 6), or UTP plus CTP plus GTP (lane 7). The sample in lane 7 was subjected to an additional round of G-50 Quick Spin column purification in the spin column buffer in which 10 mM $MgCl_2$ was replaced by 10 mM EDTA. (*B*) The hexamer ternary complex was incubated with UTP (lane 1) or with UTP in the presence of ≈4 pmol of added standard preparation RNA polymerase (RNAP; lane 2). (*C*) The hexamer ternary complex was incubated with UTP (lane 1) or with UTP and *E. coli* crude lysate (see *Materials and Methods*), which was diluted into the reaction samples to a final dilution factor of 1:1000 (lane 2) or 1:10,000 (lane 3).

RNA Cleavage Is Caused by a Transacting Factor. The spontaneous reaction of RNA cleavage in the ternary complexes was observed with the standard preparation RNA polymerase purified from *E. coli* cells. However, when the recombinant RNA polymerase prepared *in vitro* by reconstitution from individually overexpressed subunits was used in the same protocol, the cleavage of the heptamer was minimal (Fig. 1*A*, lane 2). The addition of the standard preparation polymerase to the ternary complexes formed by the recombinant enzyme enhanced the cleavage (Fig. 1*B*). Hence, the standard preparation polymerase contained a transacting cleavage factor. We were able to purify the polymerase obtained from cells from most of the cleaving activity by repeated chromatography on a Mono Q column (data not shown). Thus, the failure to cleave the heptamer was not an

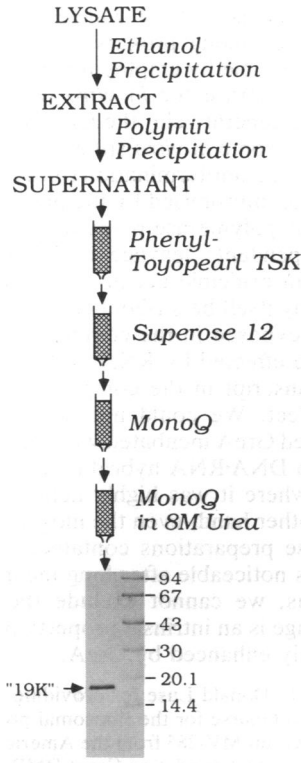

FIG. 2. Scheme of GreA purification. The final preparation of the M_r 19,000 ("19K") protein is shown as a Coomassie-stained gel band (≈3 μg of protein was loaded) next to standard molecular weight markers.

artifact of RNA polymerase reconstitution but rather reflected the lack of contamination of the recombinant enzyme with the cleaving entity.

The M_r 19,000 Cleavage Factor Is the Product of the *greA* Gene. The cleavage-enhancing activity could be detected in crude cell lysates (Fig. 1*C*). By using as the assay the conversion of CpApCpCpApCpU into CpApCpCpA, the activity was purified from the lysate to apparent homogeneity, yielding a protein with the molecular weight of ≈19,000 (Fig. 2). As the final purification step, two successive runs of the active material on a Mono Q anion-exchange column were performed under native and denaturing conditions, a procedure facilitating the removal of trace contaminants. Gel staining revealed no contaminants in the final preparation, which retained cleavage-enhancing activity at high dilution (Fig. 3). Thus, the M_r 19,000 protein was indeed the active entity.

That the M_r 19,000 protein was the *greA* gene product followed from its N-terminal amino acid sequence (determined in 25 cycles of Edman degradation), as well as from its amino acid composition, which matched the predictions from the published *greA* nucleotide sequence (1). We also obtained CNBr cleavage fragments of the M_r 19,000 protein and determined their N-terminal sequence (data not shown), yielding the total of about one-third of the whole M_r 19,000 sequence determined. All of the data fit the predictions from the *greA* gene sequence with the exception that threonine was found instead of arginine at position 85.

GreA Recycles Between RNA Polymerase Molecules. Is the transcript-cleaving activity of GreA stoichiometric or catalytic? To answer this question, we determined the kinetics of cleavage of the heptamer in ternary complexes formed by recombinant RNA polymerase by substoichiometric amounts of GreA (Fig. 3). At a molar ratio of GreA to ternary complex of 0.05:1, GreA completely cleaved the RNA heptamer after prolonged incubation, indicating that it recycled between the target complexes.

GreA Helps RNA Polymerase to Read Through an Elongation-Arresting Site. We next analyzed the effect of GreA on the elongation of RNA in the early section of the *rrnB P1* transcription unit by using the GreA-free recombinant RNA polymerase (Fig. 4). In this experiment, the purified ternary complex carrying radioactive hexanucleotide CpApCpCpApC (lane 1) was provided with only the next nucleotide, UTP (lanes 2 and 3), or with all four elongation substrates (lanes 4 and 5), in the absence (lanes 2 and 4) or in the presence (lanes 3 and 5) of GreA protein. The cleavage of the heptameric product by GreA is evident in lane 3. With the four substrates present, in the absence of GreA, about 50% of the elongating complexes stalled at a site corresponding to a transcript length of ≈12 (lane 4). This was a genuine arrested complex, which could not be chased into the full-length transcript but could be purified by size-exclusion chromatography (data not shown). Clearly, the presence of GreA protein allowed RNA polymerase to read through the arresting site (lane 5).

DISCUSSION

These experiments describe a transcription factor from *E. coli*, the GreA protein, which acts at the stage of RNA elongation. Functionally, GreA resembles the eukaryotic elongation factor SII, even though there is no substantial homology between their amino acid sequences (1, 13). Thus, the cleavage and restart of the nascent transcript appear to be a ubiquitous biochemical activity.

The biological role of the cleavage-and-restart reaction in the case of SII appears to be the release of RNA polymerase molecules that have stalled at elongation arrest sites (4, 5, 14–20). Such sites are often observed in eukaryotic *in vitro* transcription systems where they trap a fraction of the RNA polymerase molecules that pass through so that the arrested complex can neither move on nor release RNA and dissoci-

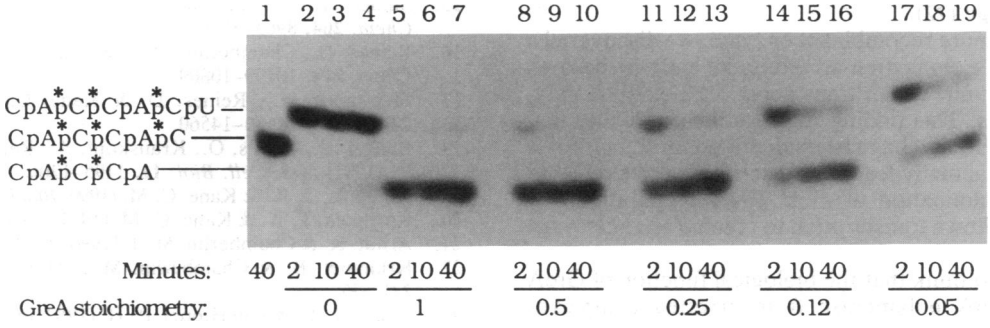

FIG. 3. Semiquantitative assay of purified GreA. Purified ternary complex formed by recombinant RNA polymerase and carrying the CpApCpCpApC hexamer was incubated under standard conditions for the indicated time periods alone (lane 1) or with nonradioactive UTP (lanes 2–19) in the absence (lanes 2–4) or in the presence (lanes 5–19) of GreA added at the molar ratios of GreA to ternary complex indicated.

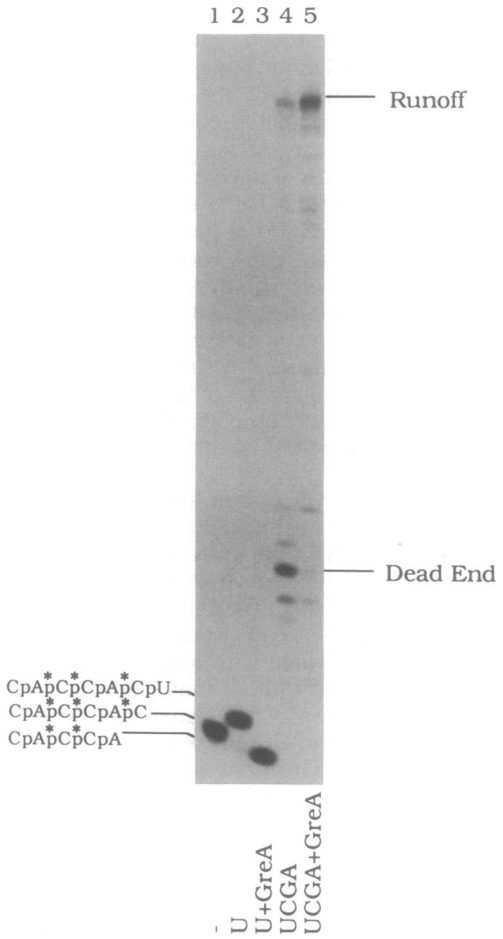

FIG. 4. Effect of GreA on the run-off transcription from the *rrnB P1* promoter. Purified ternary complex carrying the CpApCpCpApC hexamer was incubated under standard conditions alone (lane 1), with nonradioactive UTP (lanes 2 and 3), or with nonradioactive four NTPs (lanes 4 and 5) in the absence (lanes 2 and 4) or in the presence (lanes 3 and 5) of 1 molar equivalent of GreA protein.

ate. Although the molecular mechanism of this phenomenon is not clear, it appears that, in certain sequence contexts, the ternary complex has a certain probability of permanently "sinking" into a nonelongating configuration.

In prokaryotic systems, analogous transcriptional arrest of normal elongating RNA polymerases has not previously been reported. However, there have been several reports of ternary elongation complexes of *E. coli* RNA polymerase that cannot resume normal elongation after being halted for lack of a normal NTP (21, 22). These dead-end complexes are clearly analogous to eukaryotic transcription arrest complexes, since they can resume transcription elongation after transcript cleavage (22).

Using the ultrapure recombinant enzyme, we discovered a GreA-suppressible elongation-arresting site early in the *rrnB P1* transcription unit, which traps about 50% of RNA polymerase molecules. This finding suggests the possibility that some "pause sites" that have been identified for *E. coli* RNA polymerase may actually lead to transcription arrested complexes, but contamination of most RNA preparations by GreA normally allows transcription to resume after cleavage at these sites.

It is tempting to think that the biological function of GreA is to release RNA polymerase from arrested complexes formed during normal transcription. Such complexes may be responsible for the lethal effect of the *rpoB* mutation, which was originally used to identify the *greA* gene (1).

It remains to be seen how widespread the potentially arresting sites are among prokaryotic transcription units. This should be possible by employing the GreA-free recombinant RNA polymerase for *in vitro* transcription experiments. In a more general vein, the GreA phenomenon leads to realization that some *in vitro* studies of elongation, pausing, termination, and antitermination (for review, see ref. 23) could have been compromised by the presence of GreA-like activities in RNA polymerase preparations used.

The GreA-dependent cleavage of RNA in the ternary complex might in principle act in two different ways. The GreA protein may itself be a ribonuclease that attacks the 3′ terminal region exposed in the complex. Alternatively, the cleavage may be effected by RNA polymerase, perhaps by straining the transcript in the complex (3), with GreA enhancing this effect. We could not detect any nucleolytic activity in purified GreA incubated with RNA (e.g., in Fig. 4, lane 5) or with a DNA·RNA hybrid (data not shown) under the conditions where it was highly active on ternary complexes. On the other hand, even the most pure reconstituted RNA polymerase preparations contained residual cleaving activity that was noticeable after long incubation of ternary complexes. Thus, we cannot exclude the possibility that transcript cleavage is an intrinsic property of RNA polymerase, which is only enhanced by GreA.

We are grateful to Donald Luse for providing unpublished information and Richard Gourse for the ribosomal promoter. This work was supported by Grant MV-285 from the American Cancer Society and by National Science Foundation Grant DMB-9105995 under the program for U.S.–F.S.U. (former Soviet Union) Cooperative Research in Basic Science.

1. Sparkowski, J. & Das, A. (1990) *Nucleic Acids Res.* **18,** 6443.
2. Sparkowski, J. & Das, A. (1991) *J. Bacteriol.* **173,** 5256–5257.
3. Surratt, C. K., Milan, S. C. & Chamberlin, M. J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 7983–7987.
4. Reines, D. (1992) *J. Biol. Chem.* **267,** 3795–3800.
5. Izban, M. G. & Luse, D. S. (1992) *Genes Dev.* **6,** 1342–1356.
6. Burgess, R. R. & Jendrisak, J. J. (1975) *Biochemistry* **14,** 4634–4638.
7. Zalenskaya, K., Lee, J., Gujuluva, C. N., Shin, Y. K., Slutsky, M. & Goldfarb, A. (1990) *Gene* **89,** 7–12.
8. Severinov, K., Mustaev, A., Kashlev, M., Borukhov, S., Nikiforov, V. & Goldfarb, A. (1992) *J. Biol. Chem.* **267,** 12813–12819.
9. Gourse, R. L. (1988) *Nucleic Acids Res.* **16,** 9789–9809.
10. Lee, J., Kashlev, M., Borukhov, S. & Goldfarb, A. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 6018–6022.
11. Schaegger, H. & von Jagow, G. (1987) *Anal. Biochem.* **166,** 368–397.
12. Borukhov, S., Severinov, K., Kashlev, M., Lebedev, A., Bass, I., Rowland, G. C., Lim, P. P., Glass, R. E., Nikiforov, V. & Goldfarb, A. (1991) *J. Biol. Chem.* **266,** 23921–23926.
13. Hirashima, S., Hirai, H., Nakanishi, Y. & Natori, S. (1988) *J. Biol. Chem.* **263,** 3858–3863.
14. Rappaport, J., Reinberg, D., Zandomeni, R. & Weinmann, R. (1987) *J. Biol. Chem.* **262,** 5227–5232.
15. Sluder, A. E., Greenleaf, A. L. & Price, D. H. (1989) *J. Biol. Chem.* **264,** 8963–8969.
16. Reines, D., Chamberlin, M. J. & Kane, C. M. (1989) *J. Biol. Chem.* **264,** 10799–10809.
17. SivaRaman, L., Reines, D. & Kane, C. M. (1990) *J. Biol. Chem.* **265,** 14554–14560.
18. Bengal, E., Flores, O., Krauskopf, A., Reinberg, D. & Aloni, Y. (1991) *Mol. Cell. Biol.* **11,** 1195–1206.
19. Kerppola, T. K. & Kane, C. M. (1990) *Biochemistry* **29,** 269–278.
20. Kerppola, T. K. & Kane, C. M. (1991) *FASEB J.* **5,** 2833–2842.
21. Arndt, K. & Chamberlin, M. J. (1990) *J. Mol. Biol.* **211,** 79–108.
22. Krummel, B. & Chamberlin, M. J. (1992) *J. Mol. Biol.* **225,** 221–237.
23. Yager, T. D. & von Hippel, P. H. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, ed. Neidhardt, F. C. (Am. Soc. Microbiol., Washington), pp. 1241–1275.