*Proc. Natl. Acad. Sci. USA*
Vol. 77, No. 1, pp. 122–126, January 1980
Biochemistry

# Contacts between *Escherichia coli* RNA polymerase and an early promoter of phage T7

(ethylnitrosourea/phosphate triesters/dimethyl sulfate/DNA-protein interaction/gene regulation)

ULRICH SIEBENLIST* AND WALTER GILBERT

The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138

*Contributed by Walter Gilbert, September 24, 1979*

ABSTRACT **Specific contacts between the *Escherichia coli* RNA polymerase (nucleosidetriphosphate:RNA nucleotidyltransferase, EC 2.7.7.6) and the phosphates and purine bases of the A3 promoter of phage T7 cluster into three regions located approximately 10, 16, and 35 base pairs before the RNA initiation site. Two of these contain nucleotide sequences that are fairly conserved among many promoters, known as the "Pribnow box" and "−35 region" homologies; the third, just upstream from the Pribnow box, is not conserved. The polymerase binds preferentially to the coding strand and for the most part touches only one face of the DNA helix.**

The *Escherichia coli* RNA polymerase (nucleosidetriphosphate:RNA nucleotidyltransferase, EC 2.7.7.6) first binds to and then initiates transcription at DNA sites called promoters. What does the polymerase recognize? Promoter sequences exhibit some homology in two regions, the "Pribnow box" and the "−35 region"; the former located about 10, and the latter about 35, base pairs upstream from the RNA start site; their most probable sequences are T-A-T-A-A-T and T-T-G-A-C-A, respectively (1). On forming a stable association, the polymerase unwinds the DNA double helix over an 11-base-pair-long region from the middle of the Pribnow box to just past the start site (2), to expose the template for transcription. Recently, Johnsrud (3) identified contacts between the *E. coli* RNA polymerase and purines in the *lac* UV5 promoter by showing that RNA polymerase, bound to the promoter fragment, protected specific purines from, or enhanced their susceptibility to, methylation with dimethyl sulfate (4). However, these contacts are insufficient, alone, to establish the exact disposition of the polymerase on the DNA.

Here, we study close appositions between the RNA polymerase and the A3 promoter of phage T7, one of three strong promoters (5) used early in the life cycle of the phage. We examine contacts not only to the purines, but also to the phosphates along the DNA backbone. As shown by Sun and Singer (6), ethylnitrosourea preferentially ethylates phosphates. Promoter fragments bearing such phosphotriesters at positions that interfere with the binding of the polymerase can be separated from others complexed to the polymerase by filtration through nitrocellulose, which traps the protein–DNA complex (7). When the interfering ethylated phosphates are localized on the DNA backbone of the promoter fragments in the filtrate, these essential phosphate contacts give a picture of the interaction site of the polymerase. This picture is supported by the disposition of the purine contacts.

There are two ways to examine the purines. First, the polymerase can protect certain N7s of guanine and N3s of adenine against methylation by dimethyl sulfate; this determines those positions at which the polymerase is close enough to the DNA to modify the chemical attack. Second, a previously methylated guanine or adenine can interfere with the attachment of the polymerase. This determines which N7s of guanine and N3s of adenine are critical for either the formation of the complex or its maintenance.

## MATERIALS AND METHODS

**Promoter Fragment and Electrophoresis.** The 80-base-pair-long *Hha* I/*Hin*fI endonuclease fragment of T7 carries the A3 promoter and is imbedded in a larger fragment whose restriction site order is *Hpa* II, *Hha* I, *Hin*fI, *Alu* I, isolated as described (1). To label the *Hha* I terminus, we incorporate $^{32}P$ at the 5′ ends of the *Hha* I/*Alu* I fragment and then digest with *Hin*fI. Alternatively, to label the *Hin*fI terminus, we digest the *Hpa* II/*Hin*fI fragment, labeled at its 5′ ends, with *Hha* I. Thus label can be put into either strand of the *Hha* I/*Hin*fI promoter fragment. End labeling with [$^{32}P$]ATP and polynucleotide kinase, polyacrylamide gel electrophoresis, elution of DNA from gels, and autoradiography procedures are described by Maxam and Gilbert (8).

**Ethylation-Interference.** Typically, approximately 7 pmol of the 80-base-pair-long promoter fragment, labeled at one end, is suspended in 0.1 ml of 0.05 M sodium cacodylate at pH 8, then 0.1 ml of EtOH, saturated with ethylnitrosourea, is added and the solution is incubated for about 1 hr at 50°C. Next, the DNA is precipitated by the addition of 0.010 ml of 5 M ammonium acetate and 0.2 ml of EtOH and subsequently reprecipitated repeatedly from 0.1 ml of 0.5 M ammonium acetate with 2.5 vol of EtOH. Finally, the pellet is rinsed with EtOH, vacuum dried, and suspended in 0.1 ml of buffer B (10 mM Tris·HCl, pH 7.9/10 mM $MgCl_2$/100 mM KCl/1 mM dithiothreitol/0.1 mM EDTA). About 18 pmol of *E. coli* RNA polymerase (9 μg, 2.5-fold molar excess) is slowly added at 0°C and the mixture is incubated for 1–2 min at 37°C before 0.5 ml of prewarmed buffer B is added and the solution is slowly pressed through a nitrocellulose filter (Schleicher and Schuell B6, presoaked in buffer B) with a syringe. The DNA is extracted from the filter with 0.5 ml of buffer X (10 mM Tris·HCl, pH 7.5/10 mM $MgCl_2$/0.1 mM EDTA/0.5 M ammonium acetate/1% sodium dodecyl sulfate) and precipitated with 2 μg of tRNA and 2.5 vol of EtOH. After addition of 2 μg of tRNA and ammonium acetate to 0.5 M, the filtrate DNA is precipitated with 2.5 vol of EtOH. Both the filtrate and the filter-bound sample are reprecipitated from 0.2 ml of buffer R (0.3 M so-

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: Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20205.

dium acetate/5 mM EDTA), rinsed with EtOH, dried under vacuum, and then suspended in 0.015 ml of buffer E (10 mM $NaPO_4$, pH 7/1 mM EDTA). Each sample is then incubated with 0.0025 ml of 1 M NaOH at 90°C for 30 min, sealed in a capillary. Upon addition of 0.015 ml of a saturated urea solution containing tracking dyes, the samples are electrophoresed on 20% or 12% polyacrylamide/7 M urea gels (two separate loadings). A control sample (about 3 pmol) is obtained by omitting polymerase and filtration. The preparation of size markers is described under *Methylation-Protection*.

**Methylation-Protection.** About 5 pmol of the *Hha* I/*Hin*fI promoter fragment, labeled at one end only, is suspended in 0.1 ml of buffer C (50 mM sodium cacodylate, pH 8/10 mM $MgCl_2$/0.1 mM EDTA/1 mM dithiothreitol/50–150 mM KCl). Approximately 18 pmol of RNA polymerase (9 μg, about 3.5-fold molar excess) is slowly added at 0°C and then the mixture is incubated for 1–2 min at 37°C. Subsequently, 0.001 ml of 10.7 M dimethyl sulfate is added, and after 1 min at room temperature the mixture is diluted with 1 ml of prewarmed buffer B (see *Ethylation-Interference*) containing 10 mM 2-mercaptoethanol. This solution is immediately filtered through nitrocellulose (see *Ethylation-Interference*) with suction and the filter is washed once with 0.5 ml of buffer B. The filter-bound DNA is extracted and precipitated with EtOH (see *Ethylation-Interference*). The control sample (about 3 pmol of promoter fragments) is methylated in 0.1 ml as above, but without RNA polymerase, and the reaction is terminated with 0.025 ml of stop solution (8) (1 M 2-mercaptoethanol/0.5 M Tris·HCl, pH 7.5/1.5 M sodium acetate, 10 mM $MgCl_2$/0.1 mM EDTA) containing 2 μg of tRNA, and the DNA is precipitated with 2.5 vol of EtOH. Both the filter-bound and the control sample are reprecipitated from 0.2 ml of buffer R (see *Ethylation-Interference*) with 0.5 ml of EtOH, rinsed with EtOH, vacuum dried, and then treated according to the "G greater than A" method of Maxam and Gilbert (8). Each sample (final volume of 0.03 ml) is sufficient for two electrophoretic runs.

**Methylation-Interference.** About 7 pmol of promoter fragments, labeled at one end, is methylated with dimethyl sulfate, precipitated and reprecipitated several times from buffer R, exactly as the control sample in *Methylation-Protection*, except no tRNA is added at this stage. The methylated DNA is then resuspended, complexed with polymerase, filtered, and precipitated exactly as the ethylated DNA in *Ethylation-Interference*. The G greater than A method is subsequently applied to the filter-bound and the filtrate sample. A control sample is obtained by omitting polymerase addition and filtration.

**Enzymes.** *E. coli* RNA polymerase holoenzyme was purified according to Burgess and Jendrisak (9), with a modification described by Lowe *et al.* (10). The polymerase is saturated with σ subunit and is more than 95% pure, as judged from sodium dodecyl sulfate/polyacrylamide gels. Restriction enzymes and enzymes necessary for end-labeling have been described (1).

**Quantitation.** The x-ray films were scanned with an Ortec model 4310 densitometer. Badly resolved bands and weak bands are more difficult to quantitate. Therefore, only strongly interfering bands and clearly protected or enhanced bands, observed repeatedly, are recorded as contacts. Because the control sample is methylated separately from the filter-bound sample in the methylation-protection experiment, the extent of methylation may differ, producing a gradient between the samples. This problem is minimized by methylating lightly. Another problem is the overall intensity of control vs. filter-bound bands—the amount retrived after filtration varies with experiment. If any doubt exists about the base line (positions with methylation unperturbed by the polymerase), the methylation-protection experiment should be performed without filtration to establish which bands are clearly protected or enhanced (although this reduces the degree of protection or enhancement, the extent of methylation and the amount of DNA retrieved is controlled).

## RESULTS

### Polymerase–phosphate contacts

We incubated the RNA polymerase for 1–2 min at 37°C with a labeled, ethylated, T7 A3 promoter fragment: an 80-base-pair-long *Hha* I/*Hin*fI restriction piece, bearing the $^{32}P$ label at the 5′ end of one or the other of the DNA strands. During gentle filtration of this mixture through nitrocellulose the polymerase and any complexed DNA bound to the filter, while the free DNA passed through (7). The filtrate, ideally, contained only those promoter molecules carrying ethylated phosphates that had interfered with the polymerase binding, because there was only about one ethylated phosphate on each DNA fragment. However, enough RNA polymerase had to be used to saturate all the "good" promoter molecules and trap them on the filter, while not trapping the "blocked" promoters by nonspecific binding. Gentle filtration was essential, as otherwise too many "good" promoter molecules would have contaminated the filtrate; generally more than 80% of the input radioactivity remained on the filter.

To locate the positions of the interfering ethylated phosphates on the promoter DNA, we broke the DNA backbone next to the ethylated phosphate. The phosphotriester bonds are more alkali labile than the phosphodiester ones, thus alkali and heat will

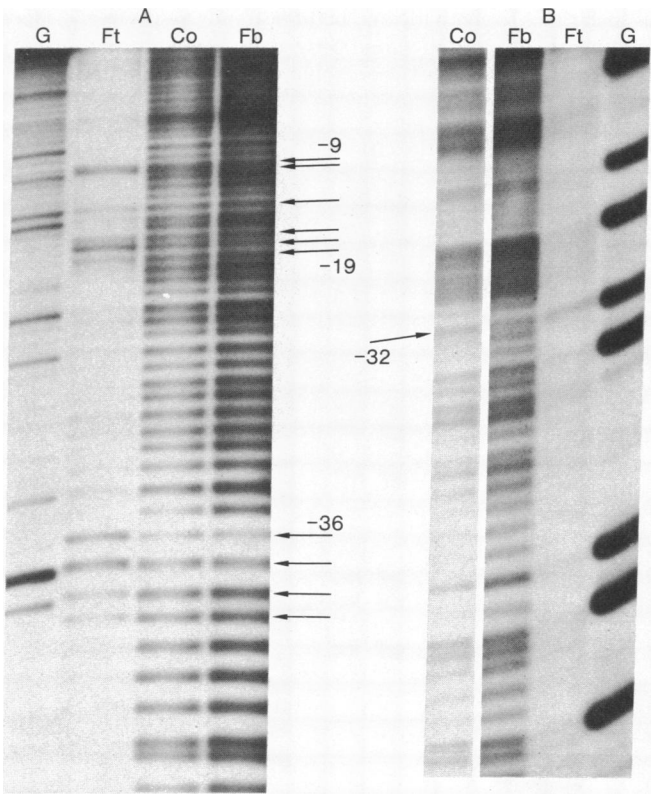


FIG. 1. Autoradiograms of the ethylation-interference experiments for the top (*A*) and the bottom (*B*) strand, with phosphate contacts marked by arrows; positions and strands are defined in Fig. 4. Fb, filter-bound sample; Ft, filtrate sample; Co, control pattern; G, pattern of Gs within the nucleotide sequence. A phosphate band comigrating with a G band represents the phosphate 5′ to that G.

create a series of broken strands, extending from the 5′ labels to the ethylated phosphate sites. Polyacrylamide gel electrophoresis under denaturing conditions separates those strands by size, and autoradiography visualizes them: a band on the x-ray film corresponds to the collection of DNA molecules that were broken at a particular distance (site) from the 5′ label.

Because the filtrate contained predominantly DNA fragments bearing interfering ethylated phosphates, the bands corresponding to these positions stand out above the background, as evidenced in the filtrate lanes, "Ft," of Fig. 1. The contacts are marked by arrows. "Fb" denotes the filter-bound pattern, and "Co" denotes the control pattern, obtained without filtration and polymerase. The sequence of Gs, determined according to Maxam and Gilbert (8) and electrophoresed in parallel here (lane "G"), serves as size markers to identify the location of the ethylated phosphate in the nucleotide sequence of the labeled strand. Fig. 1 shows the interfering ethylated phosphate positions for both the "top" (*A*), the sense, and the "bottom" (*B*), the anti-sense, strand, as defined in Fig. 4 *Upper*.

The phosphate bands trail the corresponding G bands slightly. When the DNA backbone breaks, it can do so either 5′ or 3′ to the ethylated phosphate, leaving, respectively, a 3′ OH or 3′ ethylphosphate on the labeled strand fragment. These doublets are not resolved if far enough away from the 5′ label, and they migrate slightly slower than the corresponding G bands, terminating with a 3′ phosphate.

To quantitate the effects of the blocked phosphates on the polymerase binding, we superimposed densitometer tracings of the filtrate (Ft) and the corresponding control (Co) patterns and calculated the ratio of peak heights (Ft/Co) at the position of interference. Fig. 3 *A* and *B* shows such tracings and plots of such ratios for both top and bottom strands, respectively. The amount of background varied between individual experiments, because the balance between polymerase and promoter mole-

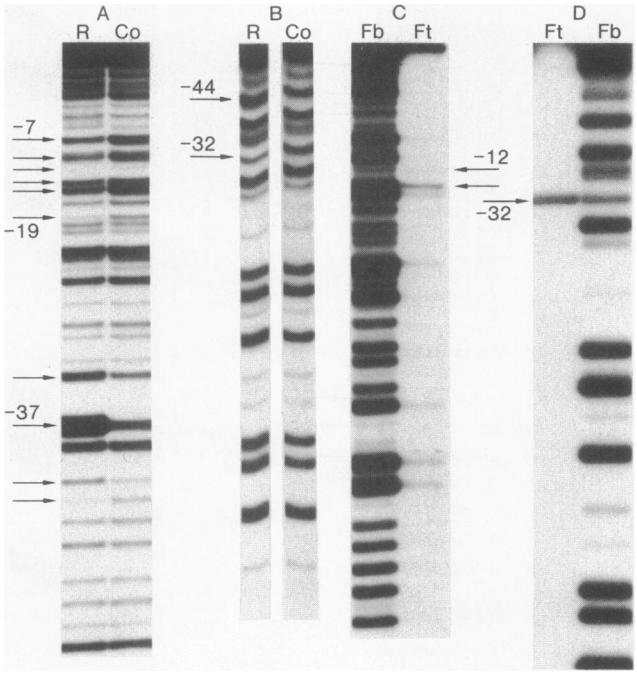


FIG. 2. Autoradiograms of methylation-protection (*A* and *B*) and methylation-interference (*C* and *D*) experiments, for the top (*A* and *C*) and the bottom (*B* and *D*) strand; purine contacts are marked by arrows. R, sample methylated in the presence of RNA polymerase; Co, control sample; Ft, filtrate sample; Fb, filter-bound sample.

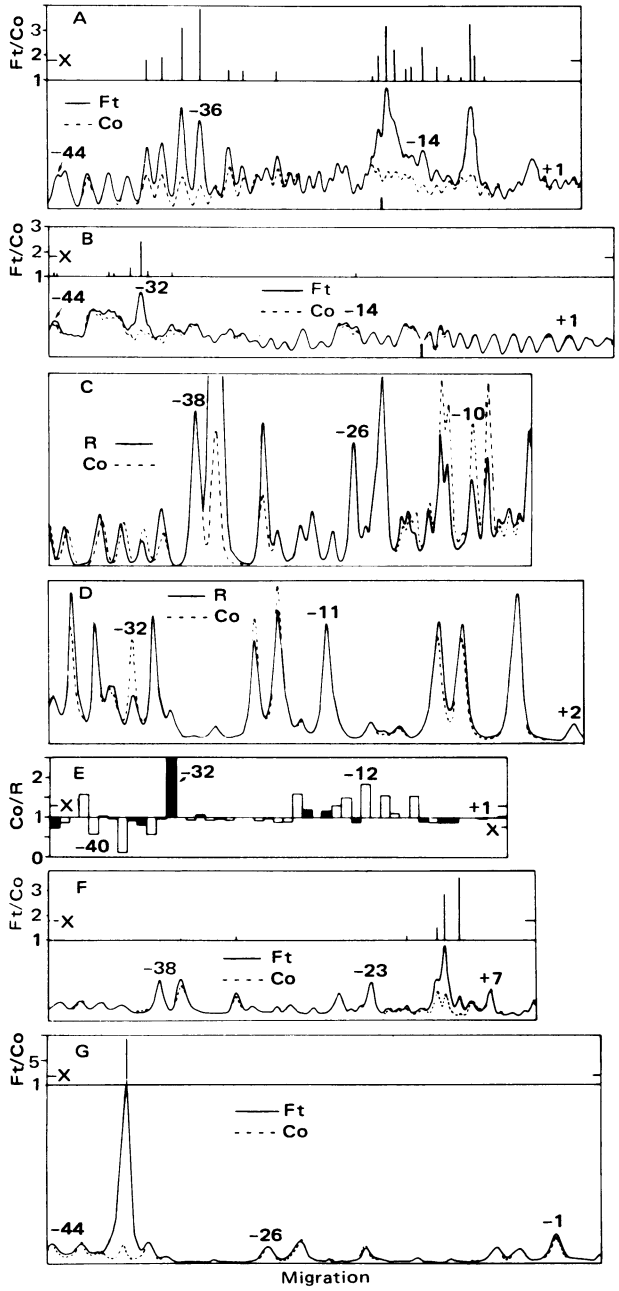


FIG. 3. Quantitation of contacts. (*A* and *B*) Superimposed densitometer tracings of control (Co) and filtrate (Ft) patterns for ethylation-interference experiments performed on the top and bottom strand, respectively. The Ft/Co peak-height ratios for the interfering ethylated phosphates are displayed above the tracings. The control tracings were adjusted to the low background band intensities in the filtrate lanes, such that Ft/Co > 1 represents a phosphate contact. (The vertical marks on the bottom of panels *A* and *B* separate tracings from two electrophoretic runs of the same and of different samples, respectively). X represents the cut-off for stronger contacts summarized in Fig. 4. (*C* and *D*) Superimposed tracings of control (Co) and RNA polymerase-affected (R) patterns for the methylation-protection experiments performed on the top and bottom strands, respectively. (*E*) Corresponding Co/R ratios, averaged over three experiments, with open bars representing purines on the top strand and solid bars representing purines on the bottom strand. The ratios are reproducible to within ± 15%. If Co/R > 1, the polymerase protected the purine from methylation and if Co/R < 1, the polymerase enhanced the susceptibility of the purine to methylation. (*F* and *G*) Superimposed tracings of control and filtrate patterns for the methylation-interference experiments performed on the top and bottom strands, respectively. Quantitation is exactly as for *A* and *B*.

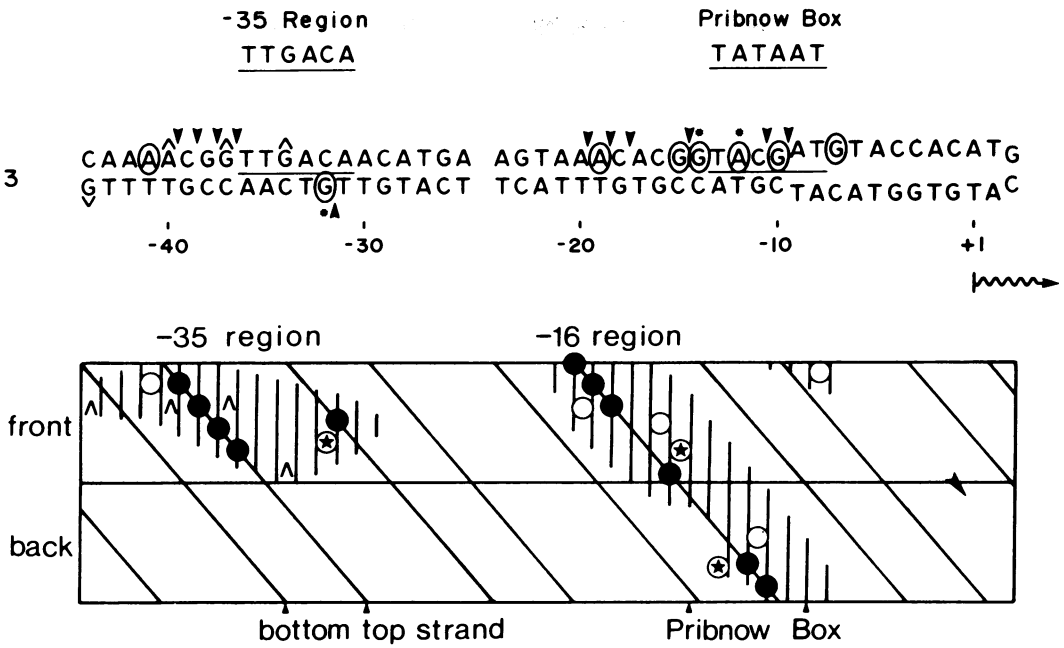


FIG. 4. (*Upper*) Sequence of the T7 A3 promoter, showing the stronger contacts to the RNA polymerase. ▼s indicate phosphate contacts; ○ and ∧ indicate purines that the polymerase protects from methylation or whose susceptibility to this methylation is enhanced, respectively; *s indicate methylated purines that interfere with polymerase binding. The most probable bases for the Pribnow box and the −35 region are shown above the corresponding regions in the A3 promoter sequence; +1 represents the start of transcription. The minimal region unwound by the polymerase is represented by a separation of the strands. (*Lower*) Planar representation of the cylindrical projection of the DNA molecule [10.5 base pairs per turn (11)], with contacts to the polymerase marked. ●s represent phosphate contacts, ★s represent methylated purines that interfere with polymerase binding, and other symbols are as in *Upper*. Contact regions, strands, front view, back view, and initiation site are indicated. Regions likely to interact with polymerase are shaded with vertical lines.

cules was not well controlled, and the degree of ethylation and the details of the filtration varied. However, repeated experiments showed the same strong contacts, defined in Fig. 3 and summarized in Fig. 4 *Upper*.

### Polymerase–purine contacts

**Methylation-Protection.** Dimethyl sulfate will methylate the N3 of A, in the minor groove, and the N7 of G, in the major groove. When the RNA polymerase is complexed to the A3 promoter fragment, the presence of the protein affects this reaction, either blocking or enhancing it. We methylated such complexes briefly at room temperature, and then, after about one purine per promoter fragment reacted, filtered the mixture quickly through nitrocellulose to trap the protein–DNA complexes. To identify those purines whose reactions had been modified, we determined the extent of methylation at each purine position. The methylated purine will depurinate on heating; alkali then can cause a $\beta$-elimination at the free sugar and create a series of broken strands extending from the 5′ labels to the site of the methylated purine, which will resolve on a denaturing polyacrylamide gel according to size. Fig. 2 shows the methylation patterns for both the top (*A*) and the bottom (*B*) strand in the presence (R) and in the absence (Co) of RNA polymerase. Positions that the polymerase protects from the dimethyl sulfate reaction, or whose susceptibility it enhances, relative to the control reaction, are marked by arrows. At position −37 there is a particularly strong enhancement, whereas the reaction at −32 shows the strongest blockage. To quantitate these contacts, we scanned the autoradiograms with a densitometer and determined the ratios of peak heights. Fig. 3 *C* and *D* shows such tracings, for the top and bottom strand, respectively, with the Co/R ratios plotted in Fig. 3*E*. Strong contacts, as defined in Fig. 3, are summarized in Fig. 4 *Upper*.

**Methylation-Interference.** A methyl group on a critical position on a purine can block the binding of the RNA polymerase to the promoter. After methylating the labeled A3 promoter fragments with dimethyl sulfate to one methylated purine per promoter molecule, we added RNA polymerase and then gently filtered the mixture through nitrocellulose, precisely as in the ethylated DNA experiment. Fig. 2 shows that bands, corresponding to positions at which methyl groups interfered, stand out above the background in the filtrate [lanes Ft of both the top (Fig. 2*C*) and the bottom (Fg. 2*D*) strand]; these contacts are marked by arrows. A methyl group at position −32 on the bottom strand blocks polymerase binding very effectively. Fig. 3 *F* and *G* shows the quantitation of the blocked purine positions, precisely as for blocked phosphate positions, for the top and the bottom strand, respectively, and Fig. 4 *Upper* summarizes the strong contacts.

## DISCUSSION

Three types of experiments identify contacts between the *E. coli* RNA polymerase and the A3 promoter of phage T7. An ethylation-interference experiment pinpoints those phosphates on the DNA backbone that, when ethylated with ethylnitrosourea, prevent the polymerase from binding to the promoter. The ethyl group may inhibit by neutralizing a negative charge, or by sterically interfering with the polymerase, or both. These phosphate contacts give a picture of the polymerase touching the "outside" of the DNA. A second experiment, methylation-protection, identifies those Gs and As that the polymerase protects from dimethyl sulfate methylation or whose susceptibility to this methylation is enhanced. The protection of the N7 of G or the N3 of A may result from steric exclusion of the dimethyl sulfate molecule by the protein or from a direct interaction with the polymerase, possibly via hydrogen bonds. The enhancements may be due to higher local concentrations of dimethyl sulfate in hydrophobic pockets formed by the protein near the purine (12); because these enhancements occur

almost exclusively next to other identified contacts (see Fig. 4 *Upper*), they signal a close protein moiety. The third experiment, methylation-interference, most likely detects those purines with which the polymerase directly interacts, those points at which a methyl group inhibits binding. Though these experiments measure different aspects of the interaction, either studying an interference with formation of the complex or examining the complex directly, they pick out identical or interspersed contacts and so identify regions critical for both polymerase recognition and binding.

As Fig. 4 *Upper* shows, the RNA polymerase binds to DNA at the two regions of strong sequence homology between promoters: the Pribnow box and the −35 region. The −35 region contacts extend out to −44; there is a region of fairly strong homology around −45 that can be detected by comparing 54 promoter sequences (unpublished). Surprisingly, the polymerase also interacts rather extensively with a third region around −16, located just upstream from the Pribnow box, where there is little homology between promoters.

Almost all known promoter mutations map in these three regions. The G at position −32 on the bottom strand is a particularly strong contact point in all three experiments; this correlates well with the fact that strong up and down promoter mutations occur at the equivalent position in other promoters (13–15).

On a three-dimensional model of DNA in the B form, all contacts upstream from the Pribnow box appear on one face of the DNA molecule; the polymerase essentially stretches along one side of the double helix. However, in the Pribnow box, as Fig. 4 *Lower* shows, the polymerase would have to touch the back of the molecule, relative to the contacts upstream. This may be illusory; because the polymerase unwinds the double helix at least as far upstream as the middle of the Pribnow box, and possibly further (2), these backside contacts could reflect the altered structure of the DNA in a transition region or the position of bases on the single strands, rather than an actual wrapping of the polymerase completely around the DNA molecule. Thus, the polymerase might initially recognize and bind all contacts upstream from the Pribnow box, lying on one face of the double helix. As the DNA strands unwind subsequently (possibly by themselves), contact points previously hidden "behind" the DNA molecule in the Pribnow box could become accessible to the polymerase. By engaging in these additional contacts, the polymerase could hold the DNA strands apart, ready for synthesis.

We thank R. B. Simpson for discussions and for *E. coli* RNA polymerase, This work was supported by National Institutes of Health Grant GM 09541-18. W.G. is an American Cancer Society Professor of Molecular Biology.

1. Siebenlist, U. (1979) *Nucleic Acids Res.* **6,** 1895–1907.
2. Siebenlist, U. (1979) *Nature (London)* **279,** 651–652.
3. Johnsrud, L. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 5314–5318.
4. Gilbert, W., Maxam, A. & Mirzabekov, A. (1976) in *Control of Ribosome Synthesis,* Alfred Benzon Symp. 9, eds. Kjeldgaard, N. O. & Maaloe, O. (Academic, New York), pp. 139–148.
5. Minkley, E. G. & Pribnow, D. (1973) *J. Mol. Biol.* **77,** 255–277.
6. Sun, L. & Singer, B. (1975) *Biochemistry* **14,** 1795–1802.
7. Hinkle, D. & Chamberlin, M. (1972) *J. Mol. Biol.* **70,** 157–185.
8. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
9. Burgess, R. & Jendrisak, J. (1975) *Biochemistry* **14,** 4364–4368.
10. Lowe, P. A., Hager, D. A. & Burgess, R. R. (1979) *Biochemistry* **18,** 1347–1352.
11. Wang, J. C. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 200–203.
12. Ogata, R. T. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 5851–5854.
13. Kleid, D., Humayun, Z., Jeffrey, A. & Ptashne, M. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 293–297.
14. Meyer, B., Maurer, R. & Ptashne, M. (1979) *J. Mol. Biol.,* in press.
15. Reznikoff, W. S. & Abelson, J. N. (1978) in *The Operon,* eds. Miller, J. H. & Reznikoff, W. S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 139–163.