Proc. Natl. Acad. Sci. USA
Vol. 74, No. 10, pp. 4266–4270, October 1977
Biochemistry

Structure of a promoter for T7 RNA polymerase*

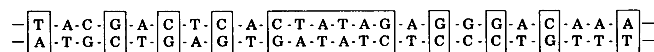
(T7 transcription/promoter DNA sequence/endoribonuclease III site/RNA nucleotidyltransferase)

JOHN L. OAKLEY AND JOSEPH E. COLEMAN

The Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510

Communicated by Frederic M. Richards, July 25, 1977

ABSTRACT We have determined the nucleotide sequence of a *Hpa* II restriction fragment of the phage T7 DNA containing a promoter for the phage-specified RNA polymerase. (*Hpa* II is a restriction endonuclease from *Haemophilus parainfluenzae*.) Mapping of the *Hpa* II restriction fragments on the T7 genome shows this promoter to be the second of tandem promoters separated by approximately 170 base pairs that begin transcription by the T7 RNA polymerase at approximately 15% of the genome. Features of the sequence involved in recognition by the T7 RNA polymerase are discussed and include the following region of hyphenated 2-fold symmetry (boxed regions are related through a 2-fold axis of symmetry at the center of the sequence shown).



This sequence includes the initiation site, since the message transcribed from this fragment begins pppG-G-G-A. Combination of our results with work of others has permitted this fragment to be mapped at the junction of T7 genes *1* and *1.1*. The RNA transcribed from this fragment begins within gene *1* and contains the RNase III cleavage site that lies between genes *1* and *1.1*. This sequence is compared to other processing sites in T7 early message.

T7 RNA polymerase (RNA nucleotidyltransferase; nucleoside triphosphate:RNA nucleotidyltransferase; EC 2.7.7.6), the product of gene *1* of bacteriophage T7, is a single polypeptide chain of molecular weight 107,000 which transcribes the rightmost 85% of the T7 genome (1, 2). At least eight promoters for the T7 enzyme exist within this region of the genome. *In vitro* transcription of whole T7 DNA by T7 RNA polymerase produces seven discrete transcripts which range in length from approximately 17,000 bases to approximately 700 bases (3). Discrete RNAs corresponding to *in vitro* transcription from the most leftward part of the genome transcribed by the T7 enzyme have not been documented since this probably results in very large RNAs, which are excluded from the gels. Hybridization to the T7 DNA of the total MRNA produced *in vitro* and *in vivo* shows that transcription does occur from at least one T7 promoter located near 15% of the genome (4). We have previously shown (5) that cutting of the T7 genome into approximately 50 double-stranded fragments with the restriction enzyme *Hpa* II (from *Haemophilus parainfluenzae*) does not destroy the promoters for T7 RNA polymerase. Transcription of the fragments results in the production of new, much shorter mRNAs. Transcription from the promoters for the T7 enzyme is highly specific; the *Escherichia coli* RNA polymerase shows no initiation at these sites. Hence determination of the nucleotide sequence of a T7 promoter should reveal significant structural features of promoter–polymerase interaction. In this paper we describe the isolation and sequence determination of a 138-base-pair restriction fragment of the T7 genome which produces *in vitro* a 54-base mRNA whose 5′ sequence corresponds to that found for normal T7 RNA transcripts from the late region.

MATERIALS AND METHODS

Materials

[γ-^{32}P]ATP and [α-^{32}P]GTP were purchased from New England Nuclear. T7 d 14 DNA, T7 RNA polymerase, and *Hpa* II restriction endonuclease were prepared as described (5). *Hae* III restriction endonuclease (from *H. aegypticus*) was purchased from New England Biolabs. Bacterial alkaline phosphatase was the gift of J. F. Chlebowski. T4 polynucleotide kinase was a gift of J. A. Steitz.

Methods

***In Vitro* Transcription of Restriction Fragments.** Restriction fragments were eluted from polyacrylamide gels by a described method (6). Transcription of these fragments by T7 RNA polymerase was carried out in a standard assay mix (5) containing [α-^{32}P]GTP.

DNA Sequencing. The nucleotide sequences of restriction fragments were determined according to Maxam and Gilbert (6).

RNA Fingerprints. RNA labeled *in vitro* with [α-^{32}P]GTP was eluted from polyacrylamide gels. Pancreatic RNase digestion and two-dimensional separation of the products by electrophoresis on cellulose acetate and DEAE-paper was according to standard procedures (7).

RNase III Digestions. Digestion of R_6 with RNase III and electrophoresis of the products on polyacrylamide gels were kindly performed by J. J. Dunn under conditions designed to differentiate primary from secondary cleavages (8). Digestions were in 20 mM Tris-HCl (pH 7.9)/5 mM $MgCl_2$/200 mM NH_4Cl/0.2 mM dithiothreitol/0.1 mM EDTA for 10 min at 37°.

RESULTS

Transcription from Small Restriction Fragments of T7 Genome. The restriction endonuclease *Hpa* II cuts the T7 genome into about 50 fragments which range in size from about 2400 base pairs to 46 base pairs (Fig. 1). The smallest fragment of 46 base pairs does not show in Fig. 1. Transcription of the *Hpa* II fragments by T7 RNA polymerase results in the production of at least eight new, much shortened mRNAs (ref. 5, and unpublished observations). Only the shortest of the original

Abbreviations: *Hpa* I and II, restriction endonucleases from *Haemophilus parainfluenzae; Hae* III, restriction endonuclease from *H. aegypticus*.

* This work forms part of a dissertation submitted by J.L.O. to the Graduate School of Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

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

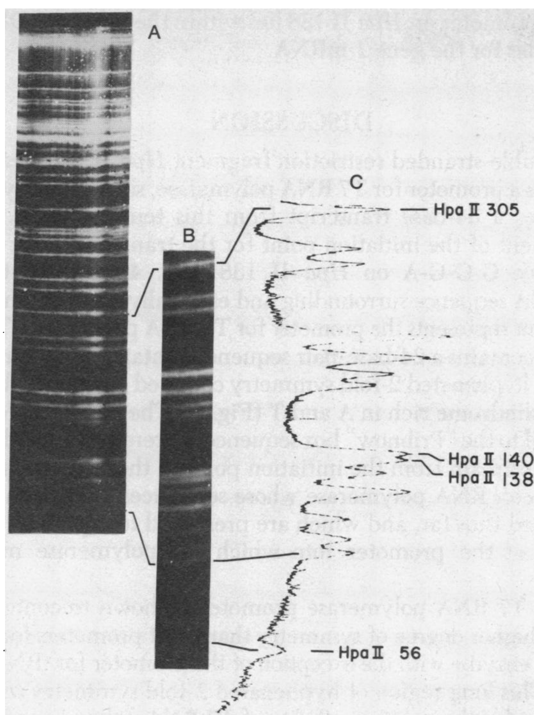


FIG. 1. *Hpa* II digest of T7 d14 DNA fractionated on (*A*) 3% polyacrylamide gel and (*B*) 5% polyacrylamide gel run in 30 mM NaH_2PO_4/40 mM Tris/1 mM EDTA and stained with ethidium bromide. (*C*) Densitometer trace of *B*.

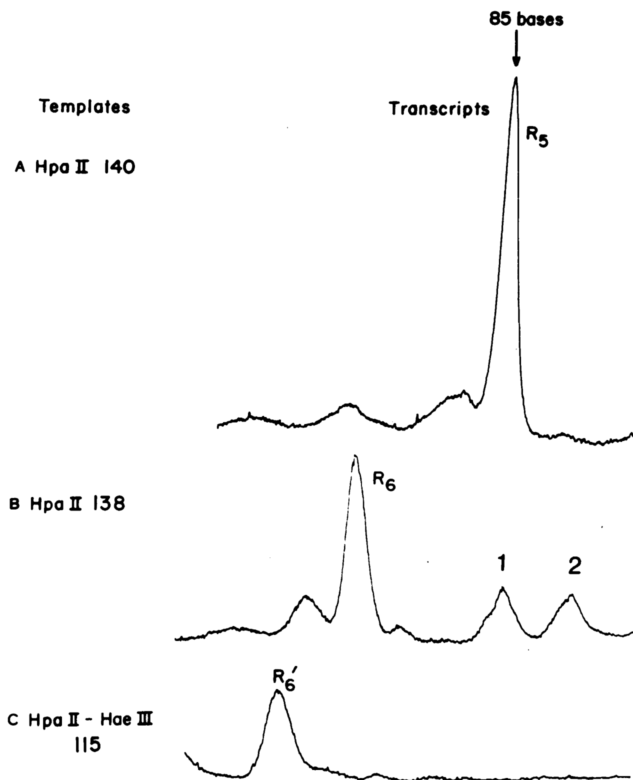


FIG. 2. RNA transcripts from restriction fragments run on denaturing 12% polyacrylamide gels in 40 mM sodium acetate, pH 5/0.1 mM EDTA/7 M urea. (*A*) R_5 transcribed from *Hpa* II 140; (*B*) R_6 transcribed from *Hpa* II 138; (*C*) R_6' transcribed from *Hpa* II-*Hae* III 115. The position of an 85-base standard (a tRNA precursor) is marked. Peak 1 in *B* is a transcript from contaminating *Hpa* II 140, while peak 2 is an end-to-end transcript of *Hpa* II 138. End-to-end transcription appears to depend on both a staggered end and the sequence in the immediate end region, since only certain *Hpa* II fragments show any evidence of it. *Hae* III cleavage of *Hpa* II 138 produces one flush end and abolishes the end-to-end transcription (*C*).

mRNAs, band VI, is unaffected by cleavage of the genome by *Hpa* II. Band VI comes from the last 3.0% of the genome and this region does not contain an *Hpa* II restriction site (5, 9). When the *Hpa* II restriction fragments of the T7 genome are separated into arbitrary size classes and these used as templates for T7 RNA polymerase, the group containing the smallest fragments (approximately 300 base pairs down to about 56 base pairs) produces the two shortest transcripts, labeled R_5 and R_6, according to the previously published nomenclature (5). By running these fragments with standard RNAs in urea gels, transcripts R_5 and R_6 were determined to be about 85 and about 50 bases long, respectively (Fig. 2). Individual double-stranded DNA fragments were eluted from the gel (Fig. 1) and tested as templates with T7 RNA polymerase. The transcripts, R_5 and R_6, come from two very closely spaced *Hpa* II fragments labeled *Hpa* II 140 and *Hpa* II 138 in Fig. 1. These two fragments were separated by prolonged gel electrophoresis and tested separately for template activity. Transcript R_5 comes from *Hpa* II 140, while transcript R_6 comes from *Hpa* II 138 (Fig. 2).

Cleavage of *Hpa* II 138 by *Hae* III. To aid in the sequence determination of *Hpa* II 138, we surveyed restriction enzymes for one that would make a further cut in *Hpa* II 138. The restriction enzyme *Hae* III makes one cut in *Hpa* II 138, removing 23 base pairs from one end. Transcription of *Hpa* II-*Hae* III 115 results in a transcript, R_6', shortened by this same number of bases (Fig. 2*C*). At the high concentrations of T7 RNA polymerase used in these experiments some end-to-end transcription of the *Hpa* II fragments takes place, as reflected by a small amount of larger bands in Fig. 2 *A* and *B*. This appears to reflect a tendency of the enzyme to pick up the staggered *Hpa* II end, as the *E. coli* RNA polymerase does to a much greater extent with many kinds of ends.

Base Sequence of *Hpa* II 138. The sequence of both strands of *Hpa* II 138 was determined by the chemical methods of Maxam and Gilbert (6). Independent sequence determinations were performed on *Hpa* II, *Hae* III 115, and on the *Hin*f fragments produced by cleavage at two sites predicted from the initial sequence (see Fig. 4). Examples of sequencing gels for two significant regions of *Hpa* II 138 are shown in Fig. 3. The DNA sequence, together with the inferred sequence of the transcript, are shown in Fig. 4. The orientation chosen is that corresponding to transcription of R_6 from left to right, since transcription by T7 RNA polymerase has been shown to be exclusively rightward (1).

Initiation Point and Sequence of R_6 mRNA. The initiation point for R_6 has been placed 54 bases from the right end of *Hpa* II 138 for the following reasons. First, the position is compatible with the estimated lengths of R_6 and R_6' relative to the *Hae* III restriction site G-G-C-C (see Figs. 2 and 4). Second, the sequence of the 5′ end of four of the *in vitro* late T7 messages have been determined by E. G. Niles (personal communication) and all begin with the common sequence pppG-G-G-A. This sequence occurs once in *Hpa* 138, beginning 54 bases from the righthand end. Third, a pancreatic RNase fingerprint of R_6 labeled by [α-^{32}P]GTP (not shown) is as would be predicted from the DNA sequence of the template. Fourth, if the enzyme, using *Hpa* II 138 as template, is given only GTP as substrate, the products are a large amount of pppG-G and lesser amounts of pppG-G-G.

Location of *Hpa* II 138. F. W. Studier and M. Rosenberg (personal communication) have mapped the fragments pro-

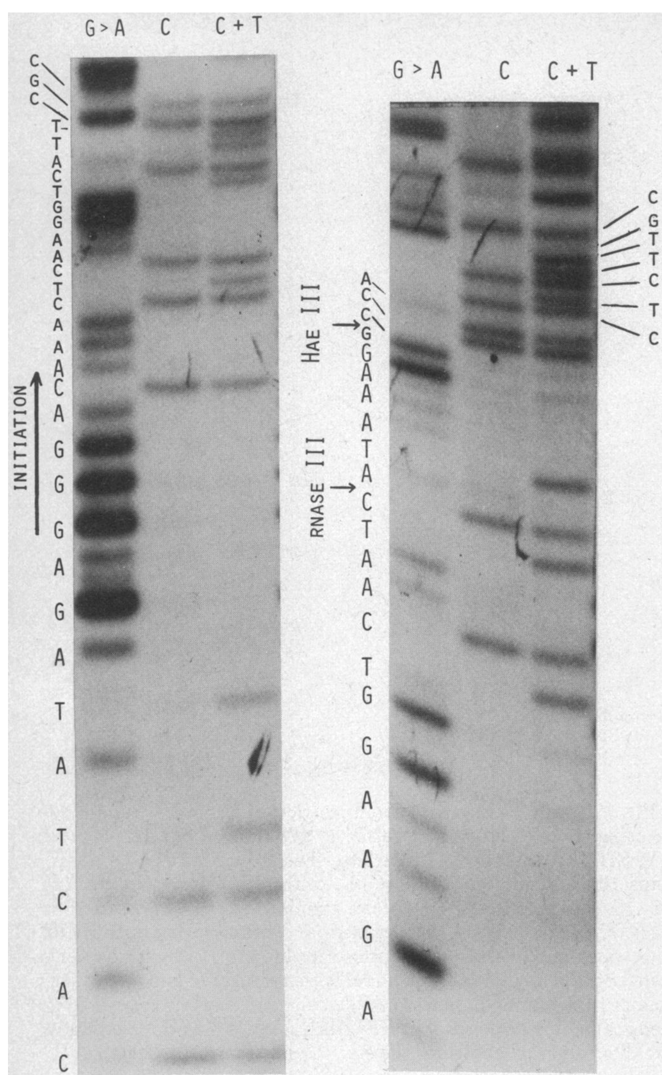


FIG. 3. An example of a sequence gel. DNA samples were prepared and the gel was run according to Maxam and Gilbert (6). While G > A is shown, the purine sequences were also checked with the A > G reaction (6). (*Left*) Sequence in the initiation region of *Hpa* II 138; *l* strand shown. (*Right*) Sequence near the *Hae* III site and the region complementary to the RNase III site in the transcript; *r* strand shown.

duced from the early region of T7 DNA by several restriction endonucleases, including *Hpa* II and *Hae* III. These restriction cleavages have also been positioned relative to genetic signals within the early region, including the RNase III cleavage sites that are known to generate the early T7 mRNAs (10). *Hpa* II 138 corresponds to a fragment mapped by Studier and Rosenberg at the junction of genes *1* and *1.1*. This fragment contains the *Hae* III cleavage site just to the left of the region coding for the RNase III cleavage site that separates the gene *1* and *1.1* mRNAs. These findings predict that the R_6 mRNA from *Hpa* II 138 should begin within gene *1* (the gene for the T7 polymerase) and might contain the complete RNase III cleavage site that lies between the gene *1* and *1.1* mRNAs. In collaboration with J. J. Dunn, we have confirmed these predictions by showing that labeled R_6 is cut by purified RNase III under conditions that are stringent for primary cleavage sites (8). The sizes of the RNA fragments produced are consistent with the 38- and 16-base fragments expected from the predicted site of cleavage (see Fig. 5 and *Discussion*). Thus it seems clear that the T7 promoter on *Hpa* II 138 lies within the region of DNA that codes for the gene *1* mRNA.

DISCUSSION

The double-stranded restriction fragment *Hpa* II 138 clearly contains a promoter for T7 RNA polymerase, since the enzyme produces a 54-base transcript from this template (Fig. 2). Placement of the initiation point for the transcript R_6 at the sequence G-G-G-A on *Hpa* II 138 (Fig. 4) suggests that the DNA sequence surrounding and especially upstream from this point represents the promoter for T7 RNA polymerase. This region contains a 24-base-pair sequence containing a high degree of hyphenated 2-fold symmetry centered around a 6-base pair palindrome rich in A and T (Fig. 4). The latter may correspond to the "Pribnow" box sequence centered about 10 base pairs upstream from the initiation point in the promoters for the *E. coli* RNA polymerase whose sequences have been determined thus far, and which are presumed to represent that region of the promoter into which the polymerase melts (11).[†]

The T7 RNA polymerase promoter is shown to contain a much higher degree of symmetry than most promoters for the *E. coli* enzyme with the exception of the promoter for $tRNA^{Tyr}$ (12). This long region of hyphenated 2-fold symmetry might account for the great specificity of T7 RNA polymerase initiation. The asymmetry that is present may provide a mechanism for asymmetric recognition of this sequence by the enzyme accounting for transcription of only one of the viral strands.

It has been suggested that regions of DNA sequence just upstream from the final binding site for *E. coli* RNA polymerase influence the ability of the enzyme to form the functional "open" complex (13). This postulate is based on the observation that once the bacterial polymerase is dissociated from a fragment of DNA it has protected from nuclease digestion, the enzyme will not rebind. Thus proper formation of the "open" complex appears to require elements of the sequence outside the protected region containing the Pribnow box. In addition "down" promoter mutations in *lac* and lambda transcription by *E. coli* RNA polymerase have been found in the region 30 to 40 base pairs upstream from the initiation point. Restriction enzymes that cut upstream from several promoters have also been shown to inactivate the promoters (13). In the corresponding region of *Hpa* II 138, centered at 33 base pairs upstream from the initiation point, there is a 27-base-pair sequence of considerable 2-fold symmetry. The two major symmetrical sequences in this region are separated by one turn of the double helix. This particular feature of symmetry in the DNA template may participate in promoter recognition.

Mapping of the Promoter on *Hpa* II 138. The restriction fragments of the T7 chromosome produced by the restriction enzymes *Dpn* II and *Hpa* I have been put into sequence by McDonell *et al.* (14). One of the *Dpn* II restriction fragments, B in the nomenclature of McDonell *et al.* (14), contains the left

[†] We have obtained evidence consistent with the suggestion that T7 RNA polymerase does melt into a short stretch of DNA at the promoter. We have been able to selectively retain promoter-containing DNA fragments on nitrocellulose filters with T7 RNA polymerase. This binding is steeply temperature dependent. Mixing of T7 RNA polymerase at low concentrations with the promoter-containing fragments *Hpa* II 138 and *Hpa* II 140 results in hyperchromia and a blue shift of the base chromophores. Addition of enzyme to non-promoter-containing fragments results in no changes at low concentrations of enzyme. At high concentrations of enzyme a hypochromia and red shift of the base chromophores occurs, probably the result of nonspecific binding.

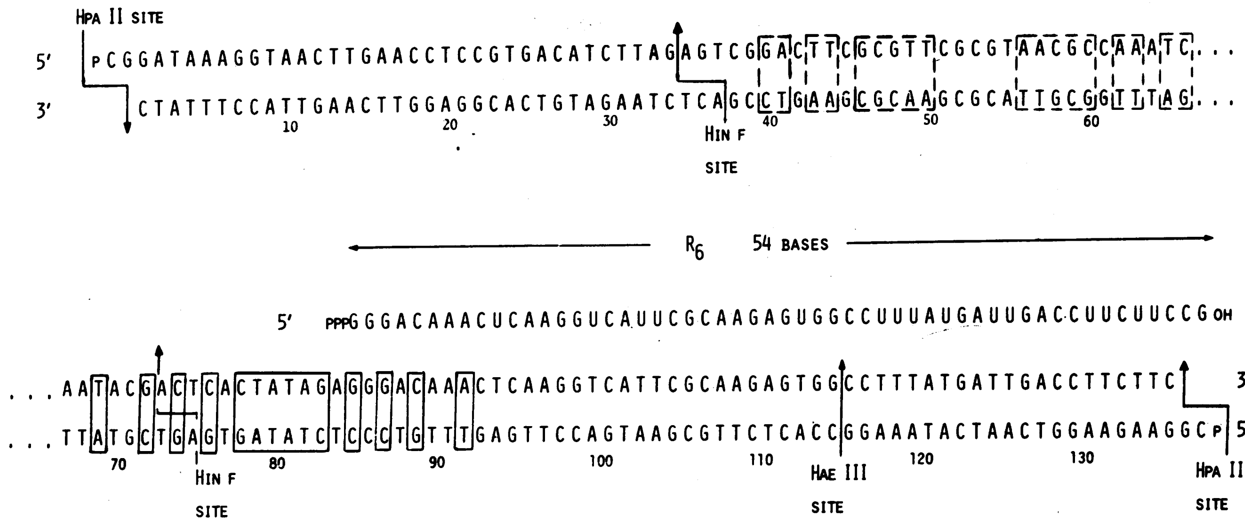


FIG. 4. DNA sequence of *Hpa* II 138 and the RNA sequence for the transcript, R_6, produced by T7 RNA polymerase.

end of the chromosome, including all of the region transcribed by the *E. coli* RNA polymerase and a small part of the region transcribed by the T7 enzyme including the first two promoters for the T7 polymerase near the end of gene *1*. Gene *1.3*, the T7 DNA ligase gene, is transcribed by both the *E. coli* and T7 RNA polymerases (15), establishing the fact that the first T7 polymerase promoter lies to the left of the terminator for the *E. coli* enzyme. The T7 enzyme reads through this terminator, since transcription by the T7 enzyme continues to the end of *Dpn* II B (J. J. Dunn and F. W. Studier, personal communication). A terminal part of gene *1* and gene *1.1* is apparently also transcribed by both polymerases. Both *Hpa* II 140 and *Hpa* II 138 come from *Dpn* II B (F. W. Studier, personal communication). The restriction fragment *Dpn* II B can be further cut into five fragments by *Hpa* I (14). One of these fragments, that labeled G in the nomenclature of McDonell *et al.* (14), produces both *Hpa* II 140 and *Hpa* II 138 when restricted further with *Hpa* II. We have confirmed that both the R_5 and R_6 transcripts are produced from the *Hpa* II digest of *Hpa* I G. Thus the promoters on *Hpa* 138 and its companion *Hpa* 140 represent the two most leftward promoters for the T7 RNA polymerase. *Hpa* II 140 has been determined to be immediately to the left of *Hpa* II 138 at approximately 14% on the chromosome (F. W. Studier, personal communication). Hence the sequence presented here is that of the second of tandem promoters which represent the beginning of the specific transcription by the phage-coded polymerase.

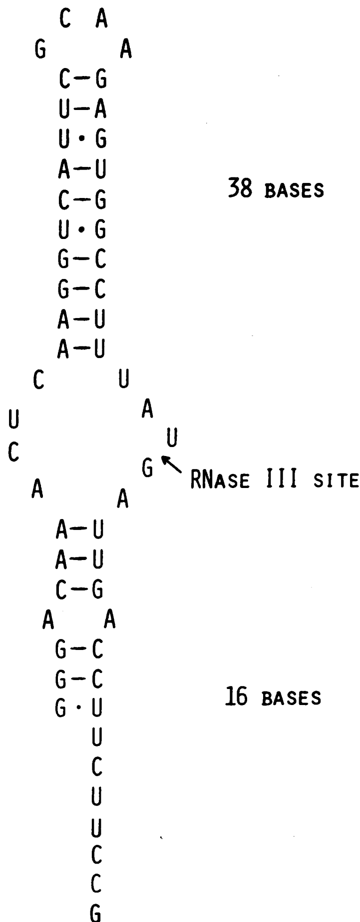


FIG. 5. RNA sequence of R_6 showing the RNase III cleavage site, including a stable base-paired loop.

The transcripts from these two most leftward T7 RNA polymerase promoters appear in amounts approximately equimolar to the messages from the other seven promoters when the *Hpa* II digest is used as template, yet transcripts initiated at these promoters do not appear prominently in gels of the *in vitro* transcripts from whole T7 DNA (3). Of the seven major *in vitro* transcripts, four have been mapped to a common terminator at 98.5% and initiate at 56, 64, 83, and 97% of the genome (3). The other three were mapped as initiating between 41 and 56% of the genome. Thus, the leftmost late region, 15–41%, did not seem to be represented in the *in vitro* transcripts. However there is some suggestion that message from these two leftmost T7 promoters may terminate to the right of gene 10, near 60% (16). Thus, the mRNA may be large enough to be excluded from the usual gels. Indeed, a large amount of radioactivity at positions corresponding to RNAs larger than bands I and II is observed when whole T7 DNA is used as template (3). This would explain the appearance of short messages from these promoters when a large part of the template distal to these promoters is removed by restriction. Initiation does occur at these promoters with the same frequency as at the seven others (ref. 5, and our unpublished results). This is perhaps a more likely explanation than postulating that some "tertiary" structure of the whole T7 DNA prevents initiation

at these promoters or that some additional factor is present *in vivo* that promotes transcription.

RNase III Processing. The early and late polycistronic messages transcribed from T7 DNA are processed *in vivo* by the *E. coli* processing enzyme RNase III (8, 16). Although these cleavages are not required for expression of T7 functions, there is some indication that cleavage stimulates the translation of certain T7 messages (17, 18). One of the RNase III processing sites for early and late messages has been mapped to the region at the right end of gene *1* (10, 16). The two promoters on *Hpa* II 140 and 138 are derived from this same region. We have shown that R_6, the transcript from *Hpa* II 138, does indeed contain an RNase III site which corresponds to the processing site between genes *1* and *1.1*. RNA sequencing for the 3′ end of the gene *1* message (19) and the 5′ end of the gene *1.1* message (20) shows that RNase III cleaves at C-C-U-U-U-A-U ↓ G-A-U. This sequence is present in R_6, and cleavage here would indeed yield the observed RNase III fragments of R_6. Thus, the R_6 sequence spans an RNase III site. There is much homology between this site and two other recently determined RNase III sites in T7. In particular, R_6 can assume a stable hairpin secondary structure by forming 10 base pairs (Fig. 5). Eighteen of these 20 base-paired residues are identical with those in the structures of other RNase III sites published by Robertson *et al.* (20) and Rosenberg and Kramer (21). In addition, the sequence of bases on either side of the loop at the base of the stem is identical with that found in the equivalent region of the site (between genes *0.3* and *0.7*) whose sequence has been determined by Rosenberg and Kramer (21). These two regions are therefore probably required for recognition by RNase III. The requirement for a second base-paired stem for recognition by the enzyme to occur is not clear, although sequence in this region may be important (19). A partially base-paired stem can be formed by R_6 (Fig. 5).

The promoter on *Hpa* II 140 is at least 205 or so base pairs (120 + 85) into gene *1*. Hence the possibility exists that the most leftward promoter for T7 RNA polymerase is within the structural gene for the enzyme. The isolation of a short double-stranded DNA of known sequence containing a functional promoter for the T7 RNA polymerase should be of considerable use in the elucidation of the molecular details of promoter-polymerase interaction and the catalytic mechanism of transcription.

We thank Dr. F. W. Studier and Dr. E. G. Niles for helpful discussions and communication of results prior to publication; Allan Maxam and Dr. Walter Gilbert for communication of their sequencing technique prior to publication; and Dr. J. J. Dunn for carrying out the RNase III digestion. We also thank Dr. J. A. Steitz and her group for helpful discussions. This work was supported by Grants GM 21919-02 and AM 09070-13 from the National Institutes of Health.

1. Chamberlin, M., McGrath, J. & Waskell, L. (1970) *Nature* **228**, 227–231.
2. Summers, W. C. & Siegel, R. B. (1970) *Nature* **228**, 1160–1162.
3. Golomb, M. & Chamberlin, M. (1974) *J. Biol. Chem.* **249**, 2858–2863.
4. Skare, J., Niles, E. G. & Summers, W. C. (1974) *Biochemistry* **13**, 3912–3916.
5. Oakley, J. L., Pascale, J. & Coleman, J. E. (1975) *Biochemistry* **14**, 4684–4691.
6. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
7. Brownlee, G. G. (1972) in *Laboratory Techniques in Biochemistry and Molecular Biology*, eds. Work, T. S. & Work, E. (Elsevier, Amsterdam), Vol. 3, pp. 1–261.
8. Dunn, J. J. (1976) *J. Biol. Chem.* **251**, 3807–3814.
9. Ludwig, R. (1975) Ph.D. dissertation, Yale University.
10. Dunn, J. J. & Studier, F. W. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 1559–1563.
11. Pribnow, D. (1975) *J. Mol. Biol.* **99**, 419–443.
12. Sekiya, T. & Khorana, H. G. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 2978–2982.
13. Gilbert, W. (1976) in *RNA Polymerase 1976*, eds. Losick, R. & Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 193–205.
14. McDonell, M. W., Simon, M. N. & Studier, F. W. (1977) *J. Mol. Biol.* **110**, 119–146.
15. Studier, F. W. (1972) *Science* **176**, 367–376.
16. Studier, F. W. (1975) *Proceedings of the Tenth FEBS Meeting*, pp. 45–53.
17. Dunn, J. J. & Studier, F. W. (1975) *J. Mol. Biol.* **99**, 487–499.
18. Hercules, K., Schweiger, M. & Sauerbier, W. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 840–844.
19. Rosenberg, M., Kramer, R. A. & Steitz, J. A. (1974) *Brookhaven Symp. Biol.* **26**, 277–285.
20. Robertson, H. D., Dickson, E. & Dunn, J. J. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 822–826.
21. Rosenberg, M. & Kramer, R. A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 984–988.