T7 Ribonucleic Acid Polymerase–Promoter Interactions†

Harry L. Osterman and Joseph E. Coleman*

ABSTRACT: The *Hpa*II restriction enzyme digest of the bacteriophage T7 genome contains among its shortest fragments two, L1 of 139 base pairs and L3 of 138 base pairs, which are specifically transcribed by T7 RNA polymerase to produce messenger ribonucleic acids (mRNAs) of 82 and 56 bases, respectively. The promoter on L3 has been shown to be the leftmost class II promoter at 14.6% of the genome. Determination of the C-terminal sequence of the T7 RNA polymerase (-Phe-Ala-COOH) by carboxypeptidase A digestion and comparison to the DNA sequence of *Hpa*II L3 show that the stop codon for gene 1, the RNA polymerase gene, is located only 26 base pairs upstream from the first T7 RNA polymerase initiation site. The promoter of L1 is a major class III promoter at 68% of the genome which begins with the message for gene 13. A ribosome binding site characteristic of class III transcripts has been identified in the message from the L1 template. Comparison of the DNA sequences of all currently sequenced class II and III T7 RNA polymerase promoters suggests a 19 base pair consensus sequence from positions −17 to +2 relative to the initiating base (+1). Base pairs −7 through −11 of this sequence constitute a *Hin*fI restriction site, and cleavage with *Hin*f abolishes all transcription. Cleavage of L1 with *Taq*I removes the left end of the 68% promoter from position −22 but has no effect on transcription from this template. From position −12 leftward all phage-specific promoters contain a relatively A-T-rich region; however, specific transcription initiation can be completely restored to the *Hin*f-inactivated promoter by religating a *Hin*f fragment of any sequence from position −12 leftward. Thus, a conserved sequence from positions +1 to −11 contains all the specific information required for initiation by T7 RNA polymerase. Using a highly purified endonuclease [specific for single-stranded regions; Strothkamp, R. E., Oakley, J. L., & Coleman, J. E. (1980) *Biochemistry 19*, 1074–1080], we have now shown that 8 base pairs (−6 to +2) of this minimum sequence are melted in the open complex of polymerase with the promoter (the sequence -ACTATAGG- in the noncoding strand). The RNA polymerase–promoter complex exposes only the noncoding strand to single-stranded endonuclease digestion; the coding strand is protected. In contrast, the *Hin*f site (−7 to −11) is completely accessible to cleavage by the *Hin*f enzyme while the polymerase is bound in the open complex. That the polymerase is bound at the adjacent position is shown by the fact that the polymerase protects the new 3′ end of the coding strand from digestion by an exonuclease after the *Hin*f cut. Protection of the coding strand from endonuclease digestion suggested that all the determinants for initiation might exist on the coding strand. The separated single strands of *Hpa*II promoter-containing fragments, however, do not initiate significant specific transcription. Reannealing the two complementary promoter strands reestabilishes normal specific mRNA formation. Hence, some feature of the double strand, most probably the sequence from −7 to −11 which constitutes the *Hin*fI site in most T7 promoters, is required for recognition and formation of a viable open complex.

The simple and yet highly specific bacteriophage T7 polymerase–promoter system presents an excellent model for the study of promoter recognition and utilization. The polymerase is a monomeric protein of molecular weight 98 092 that transcribes the late regions of the T7 genome by recognizing a unique promoter sequence

AATACGACTCACTATAG
TTATGCTGAGTGATATC

that includes the initiating base and the preceding 16 base pairs (Oakley & Coleman, 1977; Oakley et al., 1979). This 17 base pair conserved sequence is present in both class II and class III promoters, although the class III promoters exhibit a larger conserved region of 23 base pairs which includes 5 base pairs to the right in the message region and 1 base pair to the left of the above sequence (Rosa, 1979). Several class II promoters deviate from this conserved sequence at several positions (Boothroyd & Hayward, 1979; Dunn & Studier, 1981). In one case a single base change is accompanied by a reduction in transcriptional efficiency (Panayotatos & Wells, 1979). Promoter utilization by T7 RNA polymerase has also been shown to be reduced by base substitutions that alter the minor groove of the DNA helix, yet changes that occur in the major groove or deoxyribose–phosphodiester backbone have no effect (Stahl & Chamberlin, 1977). Recent experiments using a single strand specific endonuclease probe indicate that the polymerase is capable of one-dimensional diffusion along any double-stranded DNA segment, which greatly increases the probability of melting at all sequences. When a specific promoter sequence is located, the polymerase melts in to form an open complex of long lifetime in which the coding strand is protected by the polymerase, while the noncoding strand is susceptible to cutting by the nuclease (Strothkamp et al., 1980). In addition, the specific recognition of the promoter was demonstrated to involve the single free sulfhydryl group of the polymerase. Several enzymatic probes and hybrid promoters have been employed in an effort to determine the minimal sequence required for specific promoter utilization and to further define the nature of the open complex.

Materials and Methods

Materials. Restriction endonucleases *Hpa*II, *Hae*III, *Hin*fI, polynucleotide kinase, and NAD *Escherichia coli* ligase were purchased from New England Biolabs. *Taq*I was from Bethesda Research Laboratories, and bacterial alkaline phosphatase was a gift from Dr. Jan Chlebowski. T7 RNA polymerase was prepared as previously described (Oakley et al., 1975), and its concentration was determined by absorbance at 280 nm using $E_{280}^{0.1\%} = 0.74$ (Niles et al., 1974). [γ-

† From the Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510. *Received February 24, 1981.* This work was supported by National Institutes of Health Grant GM21919-05. H.L.O. was supported by National Institutes of Health Fellowship 1 F32 GM07565-01.

0006-2960/81/0420-4884$01.25/0

^{32}P]ATP, [α-^{32}P]ATP, and [α-^{32}P]GTP were purchased from New England Nuclear. Ribonucleotide triphosphates were from Sigma Chemical Co. T7 bacteriophage, strain ΔH3, and *E. coli* SY106 were used to prepare the T7 DNA as previously described (Oakley et al., 1975).

Purification and Sequence Analysis of DNA Fragments. Restriction enzyme cleavage products were fractionated by electrophoresis on polyacrylamide slab gels in TBE buffer[1] (100 mM Tris–borate and 10 mM EDTA, pH 8.3). The bands were excised from the gels, and the DNA was eluted by continued electrophoresis or by soaking (Oakley et al., 1979). Labeling of the 5′ ends of the restriction fragments and base sequence determination were according to the methods of Maxam & Gilbert (1979) with the exception of the G+A sequencing reaction which was performed at 37 °C for 25 min.

Strand Separation. The coding and noncoding strands of *Hpa*II L1 and L3 DNA fragments (Dunn & Studier, 1981) were electrophoresed on 10% strand separation gels as described by Maxam & Gilbert (1979). In order to separate any residual duplex, we recovered each strand and ran it on a second gel without denaturation by Me_2SO. Single strands were renatured by incubation in a 10-μL volume containing 0.1 M NaCl and 10 mM Tris-HCl, pH 7.4, by heating to 50 °C and cooling slowly over a period of 3 h.

Preparation of Hybrid Promoters. Several hybrid promoters were prepared by religating the following mixtures of *Hin*fI fragments: (1) *Hpa*II L2–*Hin*f 29 bp + *Hpa*II L1–*Hin*f 92 bp = 121 bp; (2) *Hpa*II L2–*Hin*f 54 bp + *Hpa*II L1–*Hin*f 92 bp = 146 bp; (3) *Hpa*II L2–*Hin*f 58 bp + *Hpa*II L1–*Hin*f 47 bp = 105 bp (see Figure 3 below for sequences of fragments). The fragments were ligated in a 40-μL volume which contained 20 mM Tris-HCl, 4 mM $MgCl_2$, 1.2 mM EDTA, 26 μL of NAD, 50 μg/mL BSA, and 5 units of *E. coli* NAD ligase for 3 h at 16 °C, pH 8.0. Following ligation the mixture was further incubated at 37 °C for 1 h with 8 units of *Hpa*II to destroy ligation at the *Hpa*II sites. The *Hin*f hybrids were then separated on a 12% TBE gel and recovered by phenol extraction and ethanol precipitation.

Transcription. Transcription of the various DNA fragments was conducted in a standard assay mixture containing [α-^{32}P]ATP or [α-^{32}P]GTP (1.3 Ci/mmol), 40 mM Tris-HCl, 30 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0.5 mM each of the four ribonucleoside triphosphates, and 2×10^{-6} M T7 RNA polymerase pH 8.0. The reactions were incubated at 37 °C for 30–120 min. Transcription was stopped by the addition of 2 M NH_4OAc, pH 5.0, containing 1 μg of tRNA, followed by ethanol precipitation. The resulting RNA pellets were dissolved in 80% formamide, bromophenol blue and xylene cyanole FF markers were added, and aliquots were loaded onto 8–12% DNA sequencing gels (see above).

Restriction Site Protection by Polymerase. The *Hin*f restriction site protection reactions contained one or more of the following: 5′-^{32}P-end-labeled *Hpa*II L1 + L3, 0.001 unit/μL *Hin*fI, 0.1 unit/μL *Taq*I, 2×10^{-6} M T7 RNA polymerase, 0.2 mM GTP, and 0.2 mM ATP in polymerase assay buffer (see Transcription). Each mixture was preincubated with the polymerase for 10 min at 37 °C, and the restriction enzyme was added for an additional 10 and 30 min for *Hin*fI and *Taq*I, respectively. The reactions were quenched by incubation with trypsin (1 mg/mL) and EDTA (0.5 mM) for 20 min; 1 μg of calf thymus DNA was added, followed by ethanol precipitation. The resulting pellets were run on DNA sequencing or nondenaturing TBE gels.

Single-Stranded Endonuclease Experiments. 5′-^{32}P-End-labeled *Hpa*II L3–*Hae*III 115 was preincubated with T7 RNA polymerase (1×10^{-6} M) for 10 min at 37 °C. The purified single-stranded endonuclease was added for an additional 10 min. Following ethanol precipitation, the reaction mixture was loaded onto a 8% sequencing gel.

Results

Gene and Protein Sequence at the C-Terminal End of T7 RNA Polymerase. In the initial studies on the restriction fragments of the T7 genome carrying promoters for the T7 RNA polymerase, this laboratory reported the isolation of three small *Hpa*II fragments (~140 base pairs long), two of which produced transcripts (56 and 82 bases long) when used as templates for T7 RNA polymerase (Oakley et al., 1975, 1979; Oakley & Coleman, 1977). With the nomenclature for the restriction fragments of the T7 genome published by Studier et al. (1979), these fragments correspond to *Hpa*II L1, L2, and L3 in the numbering system of the latter authors which we adopt here. Fragments L2 (141 base pairs) and L3 (138 base pairs) are adjacent to each other and cover 13.99–14.34% and 14.34–14.69% of the genome (Studier et al., 1979; Oakley et al., 1979). The T7 genome is transcribed unidirectionally, and this direction is referred to as left to right. Fragment L3 carries the leftmost class II promoter for the T7 RNA polymerase, while L2 is the nonpromoter-containing fragment just to the left of L3. Since L3 codes for the RNase III site defining the end of the message for gene 1 (Oakley & Coleman, 1977), it has appeared possible that fragment *Hpa*II L2 extends into the structural gene for the T7 RNA polymerase itself (Oakley et al., 1979). Hence, we determined the C-terminal residues of T7 RNA polymerase by the carboxypeptidase A method.[2] The enzyme proved relatively resistant to carboxypeptidase A, but after prolonged incubation the supernatant showed two prominent amino acids, Ala (37% yield) and Phe (14% yield), with a number of minor peaks, the most prominent of which occurred at the elution position of Gly. Translation of the messages expected from the transcription of the L2–L3 sequence in each of the three possible reading frames and in the two possible orientations of L2 relative to L3 showed that one orientation and two of the reading frames gave single stop signals in this segment. The others gave multiple stop signals. The complete base sequence for the L2–L3 segment of the T7 genome is given in Figure 1 as either the DNA sequence or that of the corresponding mRNA. Of the two reading frames which give single stop codons, one gives a C-terminal amino acid sequence of Phe-Ala-Phe-Ala which is compatible with the carboxypeptidase A digestion data. This reading frame is illustrated in Figure 1 along with the L3 sequence covering the promoter and ending at the RNase III site which defines the end of gene 1. A C-terminal Ala, a poor substrate for carboxypeptidase A, would explain the lack of rapid release of the C-terminal amino acid residues of the polymerase.

Base Pair Sequence Melted by T7 RNA Polymerase Bound at the Promoter. We have previously shown that some preparations of T7 RNA polymerase are contaminated with

[1] Abbreviations used: Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; TBE buffer, 100 mM Tris–borate, pH 8.3, and 2 mM EDTA; Me_2SO, dimethyl sulfoxide; bp, base pair; NAD, nicotinamide adenine dinucleotide; BSA, bovine serum albumin; RNA, ribonucleic acid; mRNA; messenger RNA; tRNA, transfer RNA; DNA, deoxyribonucleic acid; poly(C), poly(cytidylic acid).

[2] The carboxypeptidase A C-terminal analysis was performed by Dr. R. E. Strothkamp.

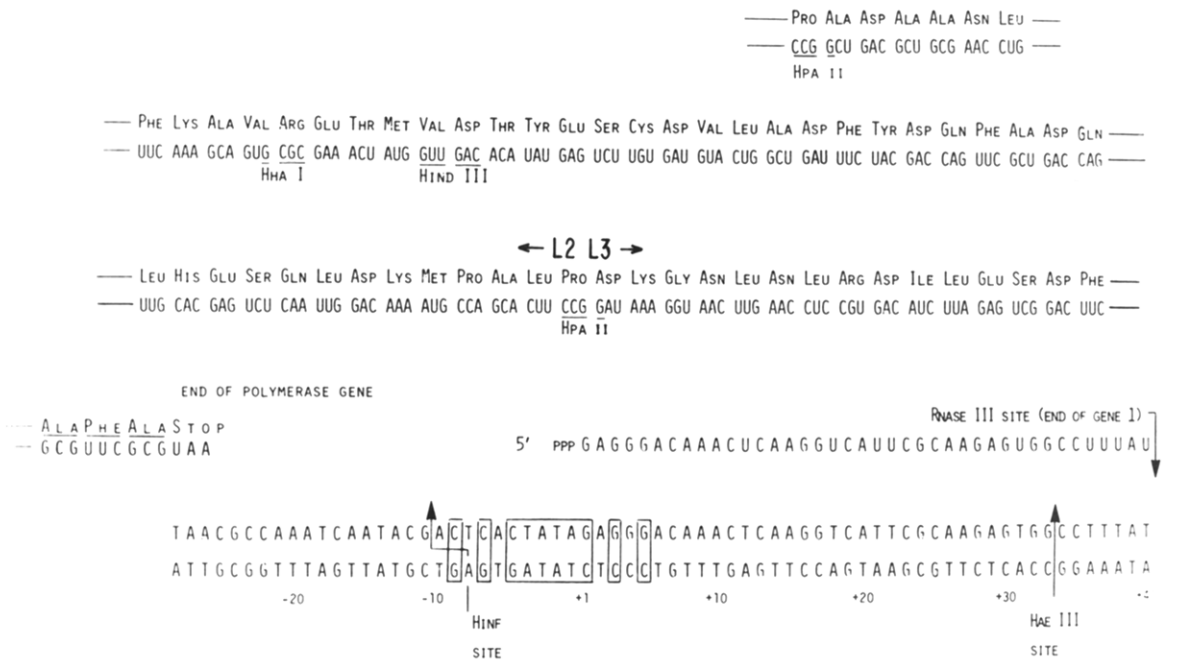

FIGURE 1: Nucleotide sequence of the C-terminal end of the T7 RNA polymerase gene (gene 1) from 13.99% to 14.69% of the T7 genome. The RNA message and the encoded amino acids for the C terminus of the polymerase are presented above the double-stranded DNA sequence of the leftmost class II promoter region contained within gene 1. The proximity of the translational stop signal for gene 1, the leftmost class II T7 RNA polymerase promoter, and the RNase III site defining the end of gene 1 are illustrated. Restriction sites are underlined or represented by arrows, the Pribnow box and accompanying hyphenated 2-fold symmetry of the promoter are outlined, and the RNase III site is marked by an arrow through the RNA.

a single-stranded endonuclease (the product of gene 3 of T7) which hydrolyzes single-stranded T7 DNA at least 10^2 times faster than the double-stranded form (Strothkamp et al., 1980). This nuclease hydrolyzes the phosphodiesters of the exposed noncoding strand at regions of the double-stranded template melted by T7 RNA polymerase. The nuclease can thus be used to probe the precise sequence melted by the polymerase when bound at the promoter (Strothkamp et al., 1980). Since the initial report, we have isolated and purified this endonuclease which will be the subject of a separate paper. The use of this highly purified endonuclease to determine the region of the promoter-containing *Hpa*II fragment L3 melted by the polymerase is shown in Figure 2. The complete L3 fragment was 5′ end labeled with [^{32}P]phosphate, and then the right-hand 23 bases were removed by restriction with *Hae*III (Figure 1). This 115 base pair template has only the 5′ end of the noncoding (upper) strand ^{32}P labeled. T7 RNA polymerase was then added to this template, followed by the purified endonuclease. Following 10 min of incubation the fragment was electrophoresed on a DNA sequencing gel, and the specific cleavage positions were determined from the radioautograph (Figure 2). A T7 RNA polymerase with negligible intrinsic endonuclease contamination was used for this experiments.

The purified endonuclease provides more flexibility in adjusting conditions, and a more detailed picture of the nuclease susceptibility of the melted region emerges. The most cleavage occurs 5′ to each of the 8 bases in the sequence -ACTATAGA-, which includes all the Pribnow box (underlined) and the first phosphodiester corresponding to the message. A subtle feature evident from these more controlled digestions is the relative protection of the phosphodiester bond 5′ to the base at position C(−5) of this sequence. Note that the base sequence (-CACTATAG-) corresponding to the bands of a sequencing gel resulting from such cleavages will be shifted one base to the left of the above sequence (Figure 2), since the labeled fragments end with the base toward the 5′ end. Attenuation

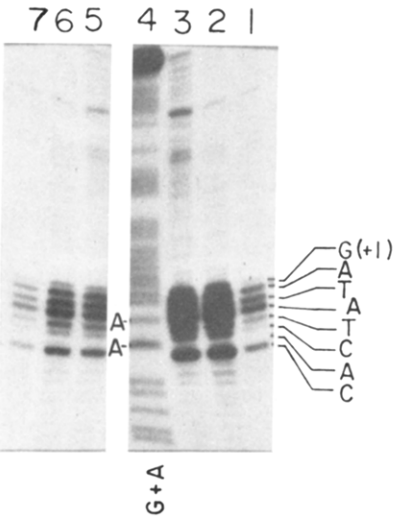

FIGURE 2: Single-stranded nuclease cleavage of the noncoding strand in the presence of T7 RNA polymerase. The radioautograph shows the ladder pattern generated by endonuclease cleavage of the melted regions of *Hpa*II L3–*Hae*III 115 (noncoding strand labeled) after preincubation with the polymerase (1–3), the Maxam and Gilbert G + A reaction (4), and a lighter exposure of 1–3 (7–5). The sequence at the right orients the cleavage pattern with respect to the G + A track. The A's labeled in the G + A track correspond to positions −7 and −3 of the promoter. See Figure 10 for locations of phosphodiester cleavages which produce these fragments.

of the conditions has made the probe more specific, and the phosphodiester bonds on the 5′ side of G(+3) and G(+4) of the message region, included in the original 10 base pair melted sequence reported by Strothkamp et al. (1980), do not appear significantly cleaved over background. The low level of background hydrolysis at every phosphodiester (much less in this purified system) is due to transient melting of all double-stranded T7 DNA by T7 RNA polymerase as previously reported (Strothkamp et al., 1980). The specific cleavages shown in Figure 2 require the presence of the T7 RNA po-

FIGURE 3: DNA sequence of *Hpa*II L1 and L2. The *Hin*fI sites are emphasized by breaks in the sequence, while all other restriction sites are underlined. (A) The sequence of *Hpa*II L1 which contains the T7 RNA polymerase promoter at 68% of the genome. The two *Hin*f fragments are designated left and right, the conserved promoter region is shown by the underlined bracket, the RNA message appears above the DNA sequence, and two potential ribosome binding sites at positions +20 and +69 of the message are shown as RNA above the DNA sequence. (B) DNA sequence of *Hpa*II L2 nonpromoter fragment from 13.99 to 14.34% of the T7 genome. The *Hin*f fragments are designated 1, 2, and 3.

lymerase as well as the nuclease and are not seen when the 5′ end of the coding strand is ^{32}P labeled (Strothkamp et al., 1980). Thus, the complementary coding strand appears to be protected from the endonuclease by binding to the active center of T7 RNA polymerase.

Promoter and Nonpromoter DNA Sequences Used To Make Hybrid T7 Promoters. The complete DNA sequence for the *Hpa*II L1 restriction fragment (139 base pairs) which carries a class III T7 RNA polymerase promoter located at 68% of the genome (Oakley et al., 1979) is given in Figure 3A. The DNA sequence for the nonpromoter fragment, L2 of 141 base pairs, is given in Figure 3B. The three fragments which result from further cleavage of L2 with *Hin*f are also indicated. Since these two *Hpa*II fragments differ in length by only two base pairs, previous sequencing strategy involved secondary cuts with *Hin*f, followed by separation and sequencing of the *Hin*f fragments. In the partial sequence of L1 previously given by Oakley et al. (1979), we misidentified the location of *Hin*f fragment 3 in L2 as coming from the left end of L1 instead of the correct fragment now shown at the left end of L1. Because of great similarity between the *Hin*f end of the two fragments, this correction involves only two base changes in the immediate promoter region, one at position −12 and one at −16, both nonessential to transcription as shown below.

Both the leftmost class II T7 RNA polymerase promoter and the class III promoter on L1 (Figure 3) contain a *Hin*f restriction site covering positions −7 through −11 (Figure 3A). Cutting with *Hin*f at this point abolishes all transcription from the promoters on L1 and L3. An example of the transcription of L1 is given in Figure 4. Positions −21 through −24 on L1 constitute a *Taq*I restriction site. In contrast to *Hin*f restriction, *Taq*I restriction, producing a fragment ending at −21, does not influence the efficiency of transcription of this fragment by T7 RNA polymerase (Figure 4).

The presence of the *Hin*f site within the essential promoter sequence and the availability of three *Hin*f fragments of different sequence from L2 provided the opportunity to test the effect on transcription initiation of altering the base sequence in the region from −12 to −23, by ligating different *Hin*f

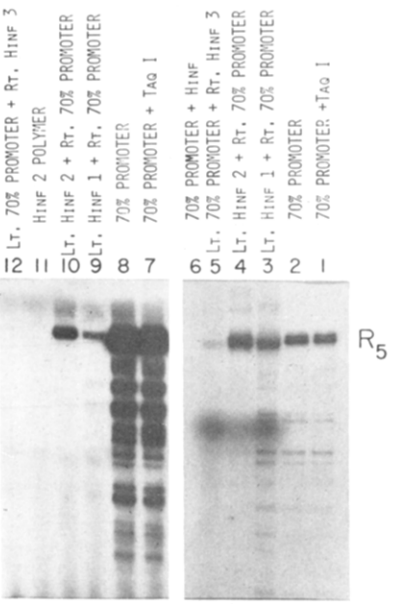

FIGURE 4: T7 RNA polymerase transcription of hybrid *Hin*f promoters. The promoters were created by ligation of fragments from *Hpa*II L1 and L2. (See Figure 3 for nomenclature and Figure 5 for sequence comparison.) The transcripts were labeled with [α-^{32}P]ATP (see Materials and Methods for details). Two separate experiments are shown in columns 1–6 and 7–12. *Hpa*II L1–*Taq* 105 (1 and 7); *Hpa*II L1 (2 and 8); *Hin*f 1 + *Hin*f right (3 and 9); *Hin*f 2 + *Hin*f right (4 and 10); *Hin*f left + *Hin*f 3 (5 and 12); polymer of *Hin*f 2 (11); L1 cleaved with *Hin*f (6).

fragments to the left end (position −10) of the class III promoter (see Materials and Methods). Radioautographs of the resulting T7 RNA polymerase transcripts using these religated fragments as templates are shown in Figure 4. Two separate experiments are illustrated in Figure 4, one of which includes as template the polymer resulting from end to end ligation of *Hin*f 2. No transcription is observed from the latter template.

Both foreign sequences substituted to the left of the original *Hin*f site restore complete specific transcription (Figure 4). The sequence variations between these hybrid promoters are

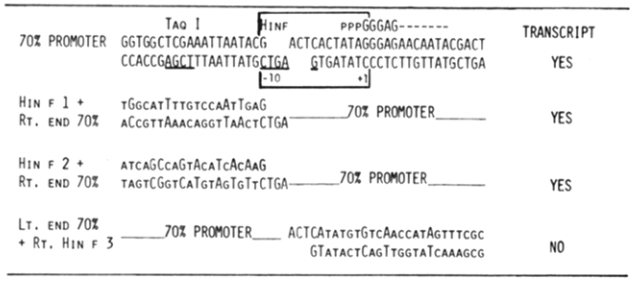

FIGURE 5: DNA sequence of *Hinf* hybrid promoter fragments. The DNA sequences of the hybrid promoters are compared to the native 70% (68%) promoter (top). The large letters represent conserved bases in the newly attached *Hinf* fragment, while small letters denote base changes. The portions of the sequence derived from the 70% promoter are not repeated. The transcription results of Figure 4 are summarized on the right. Note *Hinf* 2 will only ligate with the right end of the 68% promoter in one orientation.

summarized in Figure 5. Since all fragments have *Hinf* ends, the base changes start at position −12. In contrast, ligation of the upstream *Hinf* fragment from L1 to the right end of the middle *Hinf* fragment of L2 does not restore significant transcription, even though there happens to be a good deal of homology in the region of this hybrid corresponding to the promoter (Figure 5).

Columns 7 and 8 of the gel in Figure 4 were overloaded in order to bring out the minor mRNA fragments which are made in these transcription reactions and whose quantity depends on the template present and other less well-defined variables, since the amount of these present varies from reaction to reaction. These RNA fragments are all smaller than the authentic transcript, R_5; thus, we interpret these as representing premature termination. By means of the single-stranded endonuclease probe, T7 RNA polymerase has been observed to pause at certain sequences on the templates during transcription (Strothkamp et al., 1980). The probability of premature dissociation of the enzyme appears likely to increase at such sequences.

Reaction of Hinf with the T7 RNA Polymerase Promoter in the Presence of the Polymerase. In view of the location of the *Hinf* site at positions −7 to −11 of the promoter, just to the left of the 8 base pair sequence melted in the open complex (Figure 2), this site might be expected to be protected from *Hinf* restriction by T7 RNA polymerase binding. Hence, *Hinf* restriction was examined in the presence and absence of T7 RNA polymerase, and the radioautographs of the gel of the *Hinf* restriction products are presented in Figure 6 Details of the results are complex, but the general conclusion is clear. While T7 RNA polymerase in the open complex protects the melted region of the coding strand from single-stranded endonuclease digestion, the double-stranded *Hinf* site just to the left of the fully protected region remains accessible to *Hinf* cleavage (Figure 6).

Since separation of the promoter-containing L1 fragment (139 bp) from the nonpromoter L2 fragment (141 bp) requires extensive electrophoresis, this experiment used a mixture. L2 serves as an internal control, and no interference with protection phenomena would be expected. The set of *Hinf* fragments expected is identified in Figure 6 according to the labeling of the *Hinf* fragments given in Figure 3. Under these conditions the fragments from the two partial digestions of L2 are observed. The presence of the polymerase does not retard the cleavage by *Hinf* at the promoter site on L1. In fact, the cleavage of the nonpromoter L2 is enhanced, since the partials largely disappear. The latter is apparently a stabilization of the *Hinf* enzyme in presence of protein, since

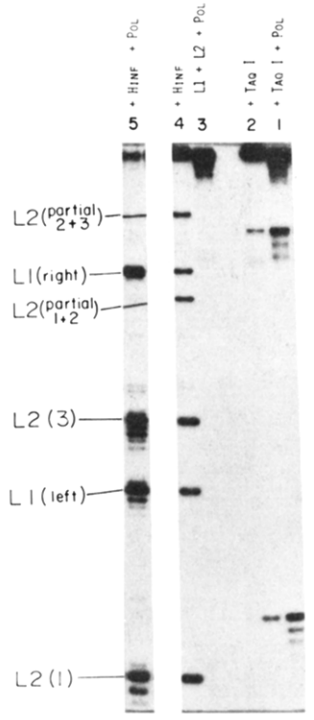

FIGURE 6: *Hinf* digestion of *Hpa*II L1 and L2 in the presence of T7 RNA polymerase. The radioautograph shows the fragments from *Hinf* cleavage of *Hpa*II L1 and L2 in the absence (4) and the presence (5) of T7 RNA polymerase, a control of *Hpa*II L1 and L2 with the polymerase (3), and *Hpa*II L1 and L2 cut by *Taq*I in the presence (1) and absence (2) of T7 RNA polymerase. The effects of the exonuclease are seen in columns 1, 3, and 5.

addition of serum albumin to the reaction mix also removes the bands due to partial cleavage.

In addition to the normal cleavage by *Hinf* in the presence of polymerase, the second unexpected finding was the cleavage of several bases from the newly generated 3′-*Hinf* ends of the restriction fragments with the exception of the 3′ end of the coding strand of the promoter, fragment L1 (right) (Figure 6, column 5). Since the strands are ^{32}P labeled at the 5′ ends, the appearance of several bands shorter than the *Hinf* fragments shows a 3′–5′-exonuclease activity to be present. As with the single-stranded endonuclease, we suspected that such a 3′–5′-exonuclease might be a low-level contaminant of the polymerase. We searched for such an activity in the preparative columns for the polymerase and found such an activity, the majority of which is removed in the flow-through of a DNA–cellulose column used to bind the single-stranded endonuclease.[3] This activity, when added to a pure *Hinf* reaction mixture, produces the digestion pattern at the 3′ ends of the *Hinf* fragments. Since the new 3′ end of the coding strand [on L1 (right)] is protected from this 3′–5′-exonucleolytic cleavage in the presence of T7 RNA polymerase (Figure 6), while the 3′ end of the noncoding strand [on L1 (left)] is not, we conclude this to be further evidence that the portion of the coding strand immediately downstream from the *Hinf* sequence is inaccessible in the open complex, as also indicated by the single-stranded endonuclease probe (Figure 2; Strothkamp et al., 1980). If *Taq*I is used to cut at position −21, clearly outside of the region of polymerase binding, then the 3′ ends of both fragments are cut by the exonuclease in the presence of polymerase, thus emphasizing the specific pro-

[3] This exonuclease is probably the *E. coli* exonuclease III which is prepared by using the same series of columns as for the T7 RNA polymerase.

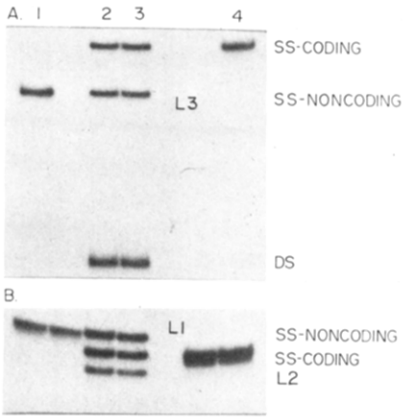

FIGURE 7: Strand separation of *Hpa*II L1 and L3. (A) A radioautograph of the strand separation according to Maxam & Gilbert (1979) of *Hpa*II L3 (2 and 3); rechromatography of coding (1) and noncoding (4) strands. The double-strand DNA appears at the bottom of the gel. (B) A radioautograph of the strand separation of a mixture of *Hpa*II L1 and L2 (2 and 3); rechromatography of the coding (4) and noncoding (1) strands of *Hpa*II L1. The band labeled L2 represents the unresolved single strands of *Hpa*II L2. The double-stranded DNA was run off this gel.

tection of the one *Hin*f end of the coding strand afforded by the T7 RNA polymerase (Figure 6, columns 1 and 2).

As observed for a number of restriction enzymes, *Hin*f will cut single strands at the specific sequence, but at a much lower rate. The cleavage rate observed on L1 in the presence of T7 RNA polymerase remains indistinguishable from that observed for double-stranded DNA; hence, the site appears to remain as a double-stranded *Hin*f in the open T7 RNA polymerase–promoter complex.

Transcription of Single-Stranded and Reannealed Promoter Fragments by T7 RNA Polymerase. In the melted open polymerase–promoter complex, the noncoding strand appears to be relatively free in solution judging from its susceptibility to single-stranded endonuclease digestion (Figure 2). This suggested that most of the determinants for specific promoter–polymerase recognition might be present on the coding strand. T7 RNA polymerase has been reported to transcribe on single-stranded poly(C) (Chamberlin & Ring, 1973). In order to determine if T7 RNA polymerase could specifically initiate and transcribe from the single-stranded sequence on the coding strand alone, we separated the strands of the two promoter-containing fragments, L1 and L3, and tested them separately as templates. Because of the necessity of having each single strand absolutely free of any of the double strand, the single strands were reelectrophoresed on a second nondenaturing gel. Radioautographs of the initial strand separation gel and the single-stranded templates are shown in Figure 7. Despite these precautions, this experiment proved to be difficult. The radioautographs of gels of a series of transcription reactions using single strands and reannealed templates are shown in Figure 8. To avoid additional processing and hence keep damage to the single-stranded DNA to a minimum, we left the 5′-^{32}P label on the template. The specific activity of the 5′-end label compared to that of the ^{32}P-labeled ribonucleoside triphosphates was not great enough to cause interference, but controls without labeled transcripts are included on the gels (Figure 8). With nucleoside triphosphates of high specific activity to ensure sensitivity, some premature termination is detected; some of this premature termination may come from transcripts initiated at the ends of the template and would account for the few bands larger than R_5 or R_6. Minor amounts of complete end to end transcription are also observed. This background does not interfere,

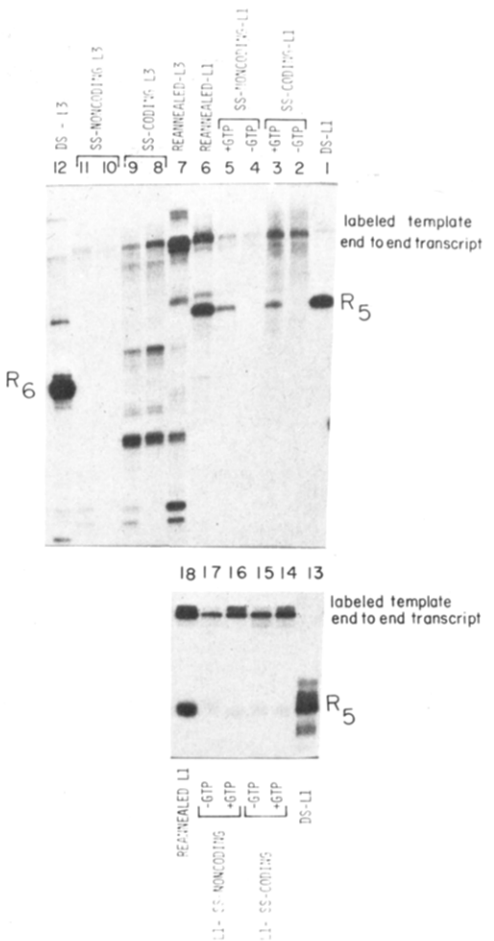

FIGURE 8: Attempt to transcribe the single strands of two promoter fragments by T7 RNA polymerase. The radioautograph shows transcription of *Hpa*II L2 templates (1–6 and 13–18) and *Hpa*II L3 templates (7–12), the characteristic R_5 (1 and 13) from double-stranded L1 and R_6 transcript (12) from L3, transcription of reannealed single strands from L1 (6 and 18) and L3 (7), L1 coding strand with (3 and 14) and without (2 and 15) [α-^{32}P]GTP, L1 noncoding with (5 and 16) and without (4 and 17) [α-^{32}P]GTP, and L3 noncoding (10 and 11) and coding (8 and 9) strands. The end to end transcripts are marked.

however, with the observation of the much greater amount of specific transcript when it is observed (Figure 8). Columns 1–6 and 13–18 show the results of two different experiments with *Hpa*II fragment L1 as template, containing the class III promoter at 68% of the genome. Both the coding and noncoding strands show an amount of specific transcript, R_5 (Oakley et al., 1979), at least 10-fold less than the amount with the reannealed double strand as template (compare columns 3, 5, and 6 and 15, 17, and 20). Reannealing of the coding and noncoding strands of L1 reestablishes transcription of R_5 (Figure 8, column 6 and 18).

We are uncertain of the origins of the light bands appearing at the positions expected for the specific transcript when either coding or noncoding single-strands are used as templates. There is the possibility that despite the apparent cleanness of the strand separation procedure (Figure 7), there remains some double-stranded contamination. On the other hand, some double-stranded structure could form at the promoter by using regions of partial complementarity on other portions of the strand. The second experiment, columns 13–18, shows considerably less of this band for the single strands as templates which may favor contamination.

When the separated strands of L3, carrying the leftmost class II promoter at 14.6% of the genome, are used as temp-

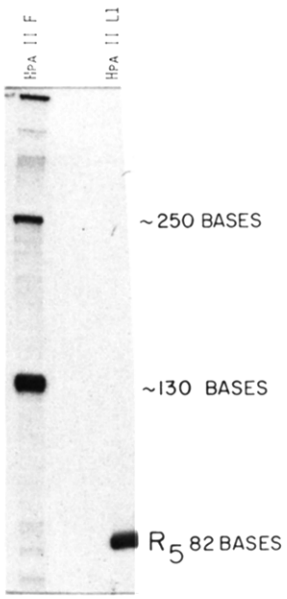

FIGURE 9: Transcription of *Hpa*II F by T7 RNA polymerase. The radioautograph shows ~130 and ~250 base transcripts from *Hpa*II F and the standard R_5 RNA of 82 bases from *Hpa*II L1.

lates, no specific transcription is observed except possibly at an extremely low level with the coding strand as template (columns 7–12, Figure 8). With the coding strand, however, a number of nonspecific transcripts are observed, although some of the bands may result from degradation of the 5′-^{32}P-end-labeled template. That secondary structure persists in the single strands of L3 is suggested by the finding that we have been unable, despite slow cooling, to reanneal the two single strands of L3 such that specific transcription is reestablished at this promoter (column 7, Figure 8). The tendency of this DNA sequence to form secondary structure has been indicated previously by the observation of several conformers of the noncoding strand on strand separation gels (Oakley et al., 1979).

Transcription by T7 RNA Polymerase from a Promoter at 1% of the Genome. The first 1.0% of the T7 genome is contained in a restriction fragment, *Hpa*II F, of 536 base pairs (Studier et al., 1979). Dunn & Studier (1981) have sequenced this fragment and identified a sequence in which 15 out of the 17 bases are the same as the sequence converved in the class II and III promoters presented here (Figures 1 and 3). If this sequence were a T7 promoter, initiation would occur at base 405 on *Hpa*II F and give a transcript of 131 bases. While no in vivo or in vitro transcript corresponding to such a promoter has been identified previously, the isolated *Hpa*II F template is transcribed in vitro by T7 RNA polymerase to give in high yield an ~130-base mRNA as predicted by the 405 start (Figure 9). This promoter deviates from the conserved 17 base pair sequence by two G → A changes at positions −11 and +1 (Dunn & Studier, 1981). Thus, this promoter does not contain a *Hin*f site. At least one *Hin*f-resistant initiation site for T7 RNA polymerase was observed when the whole *Hin*f digest of T7 DNA was transcribed in vitro (Oakley et al., 1979). The transcript would be predicted to begin with A rather than G as observed in all other T7 transcripts, unless transcription could begin with the G which occurs at position +2 of this sequence. There is some indication that T7 RNA polymerase may "wobble" at the initiation site, since GGG is produced when only GTP is used with T7 RNA polymerase and the L3 template, even though the 5′ end of the complete transcript is GAGGG (Oakley et al., 1979). It is also true that under some conditions of in vitro transcription, two

Table I

promoter	17 units[a]	nucleotide sequence		
			-17 +2	
Class III consensus sequence		gaAat	TAATACGACTCACTATAGG	GAGA
Class II	14.62	C	C A	G
	14.81	CCGG		AGA
	16.02	G G	G	C
	19.45	ACTGG	A	AG T
	19.74	ACGC	A	AGAC
	22.77	CG G	TA	GA
	27.95	AT	TGA A	
Left end	1.02	TTT	A A	
Total consensus sequence[b]		-----	tAATac-ACTCAcTA-A-g	----
T3 Promoter[c]		TTGTC	TATTTACCCTCACTAAAGG	GAAT

[a] Dunn & Studier, 1981. [b] Capital letters represent totally conserved position, lower case letters, only one change; (-) indicates the positions where two or more changes occur. [c] Adhya et al., 1981. Underlining indicates bases identical with those of the T7 promoter.

mRNAs differing in length by one base appear to be present (e.g., see Figure 4). This non-*Hin*f-containing promoter on *Hpa*II F appears to be transcribed with an efficiency comparable to the 14.6% and 68% promoters when the isolated fragment is used as template. In addition, there is a second larger transcript (~250 bases) produced from *Hpa*II F at ~10% the efficiency of the 130-base transcript (Figure 9). The origin of this transcript is not clear, since it is shorter than expected for end to end transcription of *Hpa*II F.

Discussion

The two equally efficient promoters for the T7 RNA polymerase presented here, the leftmost class II promoter (Figure 1) and the 68% class III promoter (Figure 3), both contain the identical sequence from position −16 to +1. When the known class III promoter sequences are compared as a group, a 23 base pair region from −17 to +6 is observed to be completely conserved (Rosa, 1979). Such absolute sequence conservation is not found for the reported class II promoter sequences summarized in Table I. When all are compared, there are several single base changes within this conserved region. Transcripts initiating from each of these varied class II promoters have been identified in vitro, indicating that none of these deviations, individually, eliminates promoter utilization (Kassavetis & Chamberlin, 1979; Panayotatos & Wells, 1979; Carter et al., 1981; Oakley & Coleman, 1977). The 1% promoter, reported here, is also transcribed by the T7 RNA polymerase in spite of G → A base changes at positions −11 and +1 (Dunn & Studier, 1981). In the case of the 16% promoter, the single base change C → G at position −5 within the Pribnow box was reported to reduce promoter utilization by a factor of 5 (Panayotatos & Wells, 1979). Although a complete and careful study of the relation between base changes and transcriptional efficiency has not been conducted for the class II promoters, one might expect that such sequence deviations might affect the efficiency of transcription. A similar relationship has been demonstrated for the *E. coli* RNA polymerase and the λPrm promoter, in which a single base change (λPrm^{116}) within the recognition region reduces transcription ~10-fold (Meyer et al., 1975). In the less specific *E. coli* system where two promoters of markedly different structure may have very similar activities, the relation

between promoter structure and function is not clear.

Promoters are not of course isolated sequences but are placed both spatially and temporally in an ordered genome whose transcription leads to complex and sequential functions. The T7 genome is very economical in nature. The leftmost class II T7 RNA polymerase promoter at 14.6% is located just 29 base pairs downstream from the end of gene 1, the polymerase gene itself, and only 75 base pairs upstream from a second class II promoter with an identical 17 base pair sequence at 14.81% (Panayotatos & Wells, 1979). Curiously, the RNase III site which defines the transcript of gene 1 is located between these "tandem" promoters only 23 base pairs downstream from the initiation site of the 14.6% promoter (Oakley & Coleman, 1977). Another significant feature of the DNA sequence in this region is the finding that a sequence somewhere between 14.75% and 15.0% of the genome is required for normal or primary T7 DNA replication (Tamanoi et al., 1980). The specific sequence involved in the replication appears to be an A-T-rich sequence extending rightward from position −2 of the second of these tandem promoters (Saito et al., 1980). Recent studies of in vitro replication also show the T7 RNA polymerase is required for efficient replication of T7 DNA by T7 DNA polymerase (Fischer & Hinkle, 1980; Wever et al., 1980). T7 RNA polymerase may participate in DNA replication by acting at these tandem promoters to produce a primer or by melting the DNA in this region to enable priming by the T7 gene 4 product, "primase" (Saito et al., 1980; Fischer & Hinkle, 1980).

In vivo transcription initiating from the 68% promoter produces the major T7 RNA species II which begins with the message for the gene 13 product, a coat protein. The initial portion of this transcript contains two potential Shine–Dalgarno sequences, GGAGG at position +20 and GAGGU-U at position +69 (Figure 3). Three other T7 class III ribosome binding sites for the RNA species IIIa, IV, and V have been recently identified and found to contain the common sequence of 5′-UAAGPuAGPu-3′ (Rosa, 1981). This sequence is located between 60 and 90 bases of the 5′ terminus of the RNA and in each case is immediately preceded by the nucleotides 5′-ACUU-3′ (Rosa, 1981). Of the two potential ribosome binding sites at the beginning of RNA species II, the second sequence is preceded by the bases 5′-ACUU-3′ and is located within 80 bases of the 5′ terminus. The initiator codon for this site is beyond the end of this promoter fragment but could lie within the 12 bases of the ribosome binding site as is characteristic of all other T7 examples (Dunn & Studier, 1981; Rosa, 1981).

Comparison of the T7 promoters in Table I shows the greatest sequence homology to consist of a 19 base pair region from positions −17 to +2, a sequence readily distinguished from the adjacent upstream and downstream segments which exhibit much greater sequence diversity. On this basis it can be inferred that this 19 base pair consensus sequence represents a good estimation of the minimum sequence required for specific promoter recognition and utilization by the T7 RNA polymerase. This proposal is consistent with the results that the cleavage of the 14.8% promoter by *Hpa*II (−22 to −19) (Panayotatos & Wells, 1979) and the 68% promoter by *Taq*I (−24 to −21) does not reduce the efficiency of transcription, yet cutting the promoter at the *Hin*fI site (−11 to −7) abolishes transcription as shown here for the 68% promoter and previously for a complete *Hpa*II digest of T7 DNA (Oakley et al., 1979). The availability of this restriction site within the consensus region enabled us to alter the promoter sequence beginning at position −12 and create *Hinf* hybrid promoters

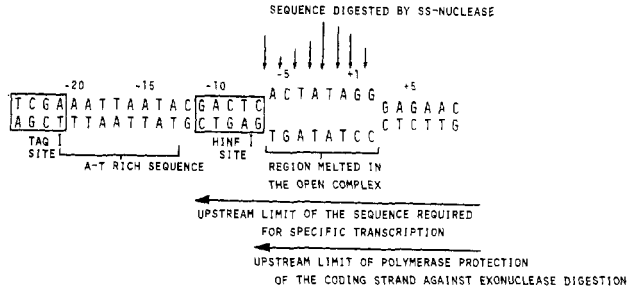

FIGURE 10: Summary of promoter probes. The DNA sequence of the 68% promoter is presented from positions −24 to +8 with respect to the initiating base. The *Taq*I and *Hin*fI sites are represented by boxes; the A-T-rich region and the region melted by the T7 RNA polymerase in the open complex are noted by brackets; the bonds cleaved by the single-stranded endonuclease are indicated by vertical arrows, and the susceptibility to cleavage of each position is indicated by the length of the arrow. The upstream limit of the polymerase protection of the coding strand from exonuclease digestion and the minimum sequence required for specific transcription are shown by horizontal arrows.

that contain a completely nonhomologous sequence in the upstream region. The specific transcription of these promoters by T7 RNA polymerase demonstrates that only 14 base pairs of the consensus sequence are required for specific initiation. In light of this result, the inability of the polymerase to transcribe the *Hinf*-digested promoter suggests that the duplex in this area is disrupted by fraying of the newly generated ends which are partially single stranded due to the nature of the staggered *Hinf* cleavage site and thus may inhibit promoter recognition. Alternately, the *Hinf* site may serve as a recognition region for the polymerase, and elimination of part of this site could destroy promoter function. The religation of a random fragment to the *Hinf* site must serve mainly to stabilize the duplex in this region and to restore the *Hinf* sequence, thus reestablishing promoter recognition.

In all known T7 promoter sequences the 10 base pair region preceding the *Hinf* site (−21 to −12) is A-T rich: 90% A-T for the class III and >70% A-T for the class II promoters. Specific transcription of the hybrid promoters occurs even when the sequence in this region is extensively altered. This A-T-rich region may influence promoter efficiency by facilitating the formation of the initiation complex and might relate to the difference in salt sensitivity of initiation at the class II promoters compared to the class III promoters (McAllister & Wu, 1978; McAllister & Carter, 1980; Carter et al., 1981).

The 14 base pair minimum sequence required for promoter recognition contains within it the 8 melted base pairs detected by single-stranded endonuclease digestion (Figure 2). The relationships of the *Taq*I site, the A-T-rich region, the *Hinf* site, and the melted sequence are summarized in Figure 10. The relative susceptibility of these diester bonds to cleavage is indicated by the length of the vertical arrows. Analysis of the individual bands in the ladder pattern generated by the purified endonuclease reveals marked differences in the susceptibility of the phosphodiester bonds in the exposed region of the noncoding strand. While all the diesters of the melted region can be cleaved by the nuclease, the region centered at the bond between C(−5) and A(−6) is relatively protected compared to the downstream end of the Pribnow box (TAG), which is most exposed. This suggests that while the noncoding strand may be relatively exposed to the solvent compared to the coding strand, the CTA sequence may still interact with the surface of the polymerase.

The melted sequence is immediately adjacent to the *Hinf* site which appears to remain double stranded in the open

complex (judging by the rate of *Hinf* cleavage) and is not protected by the polymerase binding (Figure 6). In fact, the most upstream phosphodiester of the noncoding strand in the melted sequence is again maximally exposed to nuclease digestion which suggests that the template is just emerging from the polymerase binding site at this bond. This conclusion is also consistent with the observation that while the double-stranded *Hinf* site is not blocked in the open complex, the new 3′ end of the coding strand generated by the *Hinf* cleavage remains protected from exonuclease digestion by the bound polymerase (Figure 6).

Comparison of the present data on the T7 RNA polymerase–promoter interactions with the data available on the *E. coli* RNA polymerase–promoter binding reveals some similarities to the T7 system. In the open complex of the *E. coli* enzyme, an 11 base pair sequence of the *E. coli* T7-A3 promoter is melted, including 2 base pairs of the Pribnow box and the first 2 base pairs in the message region, i.e., from position −9 to +2 (Siebenlist, 1979). Even though this result was derived by using different methods, a striking similarity can be seen in the positioning of the melted region relative to the initiating base in both open complexes.

The model for the open T7 RNA polymerase–promoter complex in which the coding strand is protected, while the noncoding strand is relatively free in solution (Strothkamp et al., 1980), suggested that the sequence of bases on the coding strand might be sufficient for specific recognition and initiation by the polymerase. The T7 RNA polymerase, however, does not appear to be able to initiate efficiently on the single coding strand as a template (Figure 9). Thus some feature of the double-stranded region must be necessary to generate the correct topology for formation of the initiation complex as is also suggested by the hybrid promoter results. The obvious candidate is the double-stranded *Hinf* sequence GACTC which appears to be essential for transcription initiation. It is significant that one of the T3 promoters which is not recognized by the T7 RNA polymerase, either in vivo or in vitro, changes the first two bases of the *Hinf* sequence to C (Adhya et al., 1981). Hence, this sequence reads -CCCTC- rather than -GACTC- (Table I). This is sufficient to abolish transcription by the T7 polymerase, since the other change, a T → A in the Pribnow box, is shared by three of the class II T7 promoters (Table I). This particular T3 promoter is located at ~1% of the genome and in contrast to the analogous T7 promoter is strongly transcribed both in vivo and in vitro by the T3 polymerase (Adhya et al., 1981). The 1% T7 promoter has the sequence AACTC; hence, it must be the purine to pyrimidine change at the first two postions of the *Hinf* site which prevents recognition by the T7 polymerase. These data strongly suggest that a partly double-stranded template is required for recognition, melting, and initiation which are the rate-limiting processes for T7 transcription (Oakley et al., 1979). The failure of T7 RNA polymerase to transcribe the single-stranded templates could also be explained by the possibility that efficient elongation requires reformation of the double-stranded template behind the transcription complex by the dissociation of the RNA–DNA hybrid and is not necessarily a reflex requirement for promoter recognition. This seems less likely, since T7 RNA polymerase has been reported to transcribe poly(C) (Chamberlin & Ring, 1973) and has been observed to move along the DNA template in response to nucleotide binding in the absence of RNA polymerization (Strothkamp et al., 1980).

Added in Proof

The total base sequence of gene 1 has been determined and shows a calculated molecular weight for T7 RNA polymerase of 98 092 (Stahl & Zinn, 1981).

Acknowledgments

We thank Dr. R. E. Strothkamp for the carboxypeptidase C-terminal analysis of the T7 RNA polymerase.

References

Adhya, S., Basu, S., Sarkar, P., & Maitra, U. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 147–151.
Boothroyd, J. C., & Hayward, R. S. (1979) *Nucleic Acids Res. 7*, 1931–1943.
Carter, A. D., Morris, C. E., & McAllister, W. T. (1981) *J. Virol. 37*, 636–642.
Chamberlin, M. J., & Ring, J. (1973) *J. Biol. Chem. 248*, 2235–2244.
Dunn, J. J., & Studier, F. W. (1981) *J. Mol. Biol. 148*, 303–330.
Fischer, H., & Hinkle, D. C. (1980) *J. Biol. Chem. 255*, 7956–7964.
Kassavetis, G. A., & Chamberlin, M. J. (1979) *J. Virol. 29*, 196–208.
Maxam, A. M., & Gilbert, W. (1979) *Methods Enzymol. 65*, 499–559.
McAllister, W. T., & Wu, H.-L. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 804–808.
McAllister, W. T., & Carter, A. D. (1980) *Nucleic Acids Res. 8*, 4821–4837.
Meyer, B., Kleid, D., & Ptashne, M. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 4785–4789.
Niles, E. G., Conlon, W., & Summers, W. C. (1974) *Biochemistry 13*, 3904–3912.
Oakley, J. L., & Coleman, J. E. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4266–4270.
Oakley, J. L., Pascale, J. A., & Coleman, J. E. (1975) *Biochemistry 14*, 4684–4691.
Oakley, J. L., Strothkamp, R. E., Sarris, A. H., & Coleman, J. E. (1979) *Biochemistry 18*, 528–537.
Panayotatos, N., & Wells, R. D. (1979) *Nature (London) 280*, 35–39.
Rosa, M. (1979) *Cell (Cambridge, Mass.) 16*, 815–825.
Rosa, M. (1981) *J. Mol. Biol.* (in press).
Saito, H., Tabor, S., Tamanoi, F., & Richardson, C. C. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 3917–3921.
Siebenlist, U. (1979) *Nature (London) 279*, 651–652.
Stahl, S. J., & Chamberlin, M. J. (1977) *J. Mol. Biol. 112*, 577–601.
Stahl, S. J., & Zinn, K. (1981) *J. Mol. Biol. 148*, 481–485.
Strothkamp, R. E., Oakley, J. L., & Coleman, J. E. (1980) *Biochemistry 19*, 1074–1080.
Studier, F. W., Rosenberg, A. H., Simon, M. N., & Dunn, J. J. (1979) *J. Mol. Biol. 135*, 917–937.
Tamanoi, F., Saito, H., & Richardson, C. C. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 2656–2660.
Wever, G. H., Fischer, H., & Hinkle, D. C. (1980) *J. Biol. Chem. 255*, 7965–7972.