JOURNAL OF VIROLOGY, July 1995, p. 4423–4430 Vol. 69, No. 7
0022-538X/95/$04.00+0

RNA Footprint Mapping of RNA Polymerase II Molecules Stalled in the Intergenic Region of Polyomavirus DNA

FRANÇOIS BRABANT AND NICHOLAS H. ACHESON*

Department of Microbiology and Immunology, McGill University, Montreal, Quebec, Canada H3A 2B4

Received 3 February 1995/Accepted 19 April 1995

RNA polymerase II molecules that transcribe the late strand of the 5.3-kb circular polyomavirus genome stall just upstream of the DNA replication origin, in a region containing multiple binding sites for polyomavirus large T antigen. Stalling of RNA polymerases depends on the presence of functional large T antigen and on the integrity of large T antigen binding site A. To gain insight into the interaction between DNA-bound large T antigen and RNA polymerase II, we mapped the position of stalled RNA polymerases by analyzing nascent RNA chains associated with these polymerases. Elongation of RNA in vitro, followed by hybridization with a nested set of DNA fragments extending progressively farther into the stalling region, allowed localization of the 3′ end of the nascent RNA to a position 5 to 10 nucleotides upstream of binding site A. Ribonuclease treatment of nascent RNAs on viral transcription complexes, followed by in vitro elongation and hybridization, allowed localization of the distal end of stalled RNA polymerases to a position 40 nucleotides upstream of binding site A. This RNA footprint shows that elongating RNA polymerases stall at a site very close to the position of DNA-bound large T antigen and that they protect approximately 30 nucleotides of nascent RNA against ribonuclease digestion.

RNA polymerase II is capable of transcribing very long stretches of DNA without interruption; for example, it can produce transcripts as long as 2×10^6 nucleotides (nt) in the case of the dystrophin gene (26). Interruption of transcription is therefore an extremely rare event in the absence of specific signals. In the cell, the template DNA is associated with histones and a variety of other proteins. It is important for efficient transcription that proteins lying in the path of RNA polymerases be displaced during elongation (50). However, several DNA-bound proteins have been shown to block elongation and/or cause termination of transcription by RNA polymerase in the intact cell. The ME1a1 protein binds within the region between the two closely linked human complement genes C2 and factor B, and plays a critical role in termination of RNA polymerase II transcribing from the upstream C2 gene (4). Proteins binding to the CCAAT box sequence upstream of the adenovirus major late promoter have been shown to mediate transcription termination by RNA polymerase II in an orientation-dependent manner (12, 13). A sequence located between the P1 and P2 promoters of the mouse c-*myc* gene directs transcription termination and is bound by a protein called TBF I (42). The exact mechanism by which these proteins interact with transcribing RNA polymerase molecules is unknown.

RNA polymerase II transcribing the late strand on polyomavirus DNA does not terminate efficiently anywhere along the 5.3-kb circular genome (1, 3, 8). A fraction of RNA polymerases can complete several traverses of the polyomavirus genome, giving rise to multi-genome-length RNAs (2, 20). However, run-on experiments detected an unusually high density of RNA polymerases on the late strand within the intergenic region of polyomavirus DNA, suggesting that RNA polymerases stall in that region in vivo (46). Bertin et al. (7) defined a 164-nt region spanning nt 173 to 11 within which these excess polymerases accumulate. This region contains the four binding sites for polyomavirus large T antigen denoted 1/2, A, B, and C (Fig. 1) (15, 17, 39). Furthermore, stalling in this region was shown to depend on the presence of functional large T antigen (7). By use of site-directed mutagenesis, it was also shown that binding of large T antigen to site A, but not to site B or C, is essential for the stalling and the accumulation of RNA polymerase II in this region (6).

The mechanisms by which polyomavirus large T antigen and other DNA-binding proteins block elongation by RNA polymerase II are not known in any detail. A mutant *Eco*RI protein unable to cleave DNA but retaining its site-specific DNA-binding properties is capable of blocking progression of several unrelated RNA polymerases in vitro (37, 38) and therefore appears to act nonspecifically. On the other hand, TTF-1, a protein involved in termination of transcription by RNA polymerase I, must interact specifically with that polymerase, because it does not block elongation by RNA polymerase II or III, *Escherichia coli* or phage T3 RNA polymerases (29).

To examine the interaction between RNA polymerase II and large T antigen bound to polyomavirus DNA, we set out to map precisely the position on the template DNA where RNA polymerases stall in vivo. Use of standard footprinting techniques was not feasible, since only about 1% of viral DNA present in cells during the late phase of infection is transcribed (9a, 33) and only a fraction of the transcribed DNA contains RNA polymerases stalled near the large T antigen binding sites (6, 7). Therefore, we chose to define the location of these RNA polymerases indirectly by analysis of the nascent RNA chains associated with them. Elongation of these RNA chains in the presence of radioactive precursors in vitro, followed by hybridization with DNA fragments extending for different lengths into the region where large T antigen binds, allowed us to locate the proximal or front end of stalled RNA polymerases (the end that comes in closest contact with bound large T antigen). Ribonuclease treatment of nascent RNAs on viral transcription complexes, followed by elongation in vitro and hybridization to DNA fragments, allowed us to locate the distal or rear end of stalled RNA polymerases (the end most distant

* Corresponding author. Mailing address: Department of Microbiology and Immunology, McGill University, 3775 University St., Montreal H3A 2B4, Quebec, Canada. Phone: (514) 398-3921. Fax: (514) 398-7052. Electronic mail address: AD13@ musica.McGill.CA.

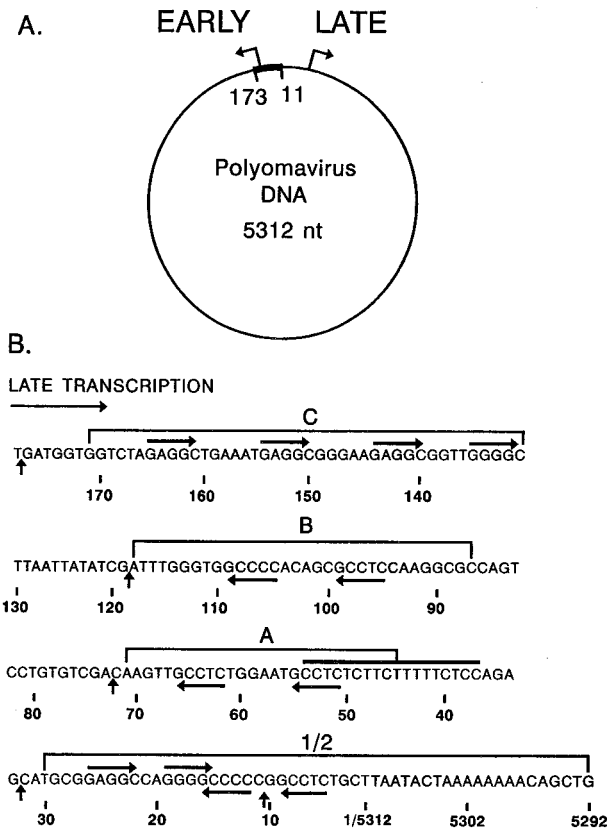

FIG. 1. Map of the stalling region on polyomavirus DNA. (A) Polyomavirus genome with the early and major late transcription initiation sites. RNA polymerase II molecules that initiate at the late promoter transcribe in a clockwise direction around the circular genome up to the intergenic region, where some polymerases stall and accumulate in vivo between nt 173 and 11. (B) Nucleotide sequence of the region in which RNA polymerases stall. Shown is the DNA strand corresponding to RNAs made by RNA polymerases that initiate transcription at the late promoter. The DNase I footprints of bound large T antigen (15) are shown by the brackets. The horizontal arrows denote the G(A/G)GC pentanucleotide sequence recognized by the large T antigen. The upward arrows indicate the positions of the 5′ ends of the DNA fragments shown in Fig. 2. The C/T-rich region where RNA polymerases pause in vitro is overlined.

from bound large T antigen). Together, this information shows that RNA polymerases are stalled just upstream of large T antigen binding site A on polyomavirus DNA.

MATERIALS AND METHODS

Virus strains and cell culture. The wild-type polyomavirus strain (AT3-Modori) used in this study was described elsewhere (6). The thermosensitive strain AT3ts25E contains a single mutation at amino acid 769 of large T antigen, introduced by site-directed mutagenesis, rendering the protein thermolabile (7). Mouse 3T6 cells were cultured as described previously (7) and were infected at a multiplicity of 1 to 10 PFU per cell.

Transcriptional run-on assay and analysis of RNA-DNA hybrids. Viral transcription complexes were isolated from infected cells 30 to 40 h after infection and incubation at 37°C or 72 h after infection and incubation at 32°C, as described previously (7, 46). Run-on assays containing viral transcription complexes from 1×10^7 to 2×10^7 cells were performed at 26°C in a 250-μl reaction mixture containing 10 mM Tris-HCl (pH 7.9); 150 mM $(NH_4)_2SO_4$; 5 mM $MgCl_2$; 2.6 mM β-mercaptoethanol; 0.2% Sarkosyl; 120 U of RNAguard (Pharmacia); 25% sucrose; 400 μM (each) UTP, ATP, and GTP; 0.2 to 1.0 μM unlabeled CTP; and 250 μCi of [α-^{32}P]CTP (0.33 μM, 3,000 Ci/mmol). The total concentration of CTP was between 0.5 and 1.3 μM in different experiments. The transcription reaction was stopped by the addition of 250 μl of phenol-chloroform (1:1), and RNA was separated from unincorporated nucleotides by exclusion chromatography on Sephadex G-50. Radiolabeled RNA was hybridized with 10 μg of single-stranded recombinant M13 phage DNA containing the viral DNA fragment of interest in 100 μl of 25% formamide–0.4 M NaCl for 4 h at 37°C.

Hybrids were assayed by collection on nitrocellulose filters, ribonuclease treatment, and counting in an LKB Rackbeta scintillation spectrometer as previously detailed (46). Alternatively, 100-μl hybridization mixtures were added to 416 μl of 10 mM Tris-HCl (pH 7.5)–0.3 M NaCl–5 mM EDTA–31 μl of ribonuclease A (10 μg/ml)–3 μl of ribonuclease T_1 (100 U/ml) (final concentration, 0.5 μg of ribonuclease A and 0.5 U of ribonuclease T_1 per ml), and then the mixtures were incubated at 30°C for 30 min to digest the unhybridized RNA. The concentration of NaCl was then brought up to 0.9 M, and the hybrids were collected on 13-mm-diameter nitrocellulose filters (Schleicher & Schuell, 0.45-μm pore size). The filters were washed twice with 50 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [pH 7])–1 mM EDTA–0.75 M NaCl, and RNA-DNA hybrids were released from the filters by digestion of single-stranded DNA in 30 mM Na acetate (pH 4.5)–0.3 M NaCl–1 mM $ZnSO_4$–5% glycerol–325 U of SI nuclease (Pharmacia) per ml at 37°C for 1 h (48). RNAs were denatured and analyzed by electrophoresis on 8% polyacrylamide–6.2 M urea gels in 89 mM Tris-borate buffer.

RESULTS

Mapping the 3′ end of nascent RNAs associated with stalled RNA polymerases. Run-on experiments indicated that there are 10- to 20-fold more RNA polymerases in region 173-11 than in regions immediately upstream (nt 406 to 178) or downstream (nt 11 to 5144) on the late strand of polyomavirus DNA (6, 7). Since accumulation of stalled RNA polymerases in this region depends both on the presence of functional large T antigen and on the integrity of binding site A, we reasoned that RNA polymerases transcribing viral DNA in vivo may cease elongation as a result of encountering large T antigen bound to site A. Large T antigen bound to site A protects nt 45 to 70 in an in vitro DNAse I footprint assay (Fig. 1B [15]); this would place the proximal end of these stalled RNA polymerases just upstream of nt 70.

If the stalled RNA polymerases accumulate at a specific site, extension of their nascent RNA chains in vitro will lead to preferential synthesis of RNA from regions that lie beyond their 3′ ends. To map the position of these 3′ ends, we hybridized radioactive in vitro run-on RNA to a series of single-stranded DNAs, cloned into M13 bacteriophage vectors, that are complementary to late RNA. These DNAs (Fig. 2A) share a common 3′ end (at nt 406), but their 5′ ends extend progressively farther into the region of interest, to nt 178, 119, 73, 33, and 11; these fragments are hereafter denoted by their 5′-terminal nucleotide. DNAs that do not extend beyond the stalling site will hybridize with only low levels of radioactive RNA, generated by the low density of RNA polymerases transcribing regions upstream of the stalling site. DNAs that extend beyond the stalling site, in contrast, will hybridize with higher levels of radioactive RNA, arising from the high density of stalled RNA polymerases. Quantitative hybridization of run-on RNAs with this series of DNA fragments can therefore define the position of the 3′ end of nascent RNA chains associated with stalled RNA polymerases.

Figure 2B shows averaged results from three independent experiments in which run-on RNA was hybridized to these DNA fragments. The transcriptional activity corresponding to each DNA was calculated by dividing the counts per minute hybridized to the fragment by the number of C residues in RNA complementary to the fragment (see reference 46), since radioactive CTP was used to label the RNA. This corrects for the variable content of C in the different fragments, so that a uniform distribution of RNA polymerases should give rise to the same transcriptional activity in all fragments. In each experiment, the transcriptional activities were normalized by division by the activity calculated for fragment 178, and the averages of these ratios are shown graphically in Fig. 2B.

Transcriptional activity in fragment 119 was not significantly different from that in fragment 178; therefore, the high density of stalled RNA polymerases must lie downstream of nt 119.

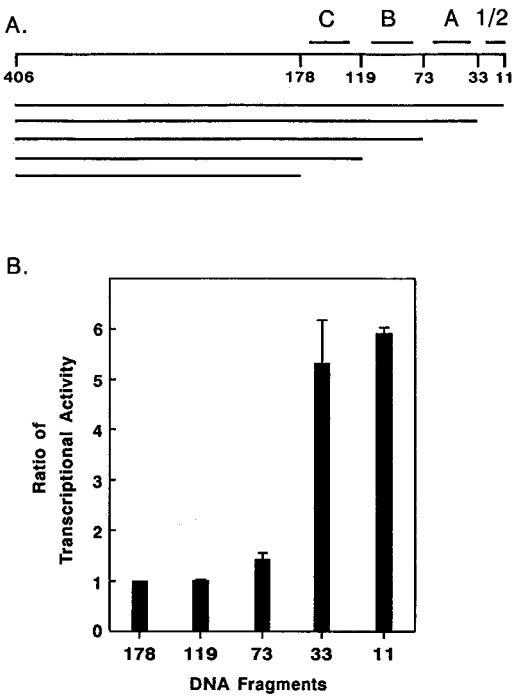

FIG. 2. Mapping the 3′ end of the nascent RNA. (A) Schematic representation of the single-stranded DNA fragments used. The five fragments have their 3′ ends at nt 406 and their 5′ ends at nt 178, 119, 73, 33, and 11, respectively. (B) RNAs labeled during 10-min run-on transcription reactions were hybridized with the set of DNA fragments described above. The ribonuclease-resistant radioactivity hybridized with each DNA fragment was determined and corrected by being divided by the number of C residues expected in RNA transcripts of each fragment. Results were normalized to the value found for fragment 178. Error bars show the standard deviations of the ratios of transcriptional activity (results from three experiments for fragments 178, 119, 73, and 33 and two experiments for fragment 11).

Transcriptional activity in fragment 73 was about 1.4-fold higher than those in fragments 178 and 119, and transcriptional activities in fragments 33 and 11 were 5- to 6-fold higher; these results suggest that the 3′ ends of RNAs associated with stalled RNA polymerases lie somewhere upstream of nt 73.

A closer analysis of these results suggests that the RNA polymerases are stalled only a very few nucleotides upstream of nt 73. The transcriptional activities (counts per minute per C residue) for fragments 178 and 119 can be taken as representing the average activity of RNA polymerases transcribing the region upstream of the stalling site. This value was multiplied by the number of C residues in each of the other fragments, and the results were subtracted from the actual counts per minute incorporated into RNA hybridized to each fragment. This calculation yields a value that represents the excess transcription of each region by stalled RNA polymerases, expressed as counts per minute incorporated into C residues. Table 1 compiles these results for one experiment and shows that the ratio of excess counts per minute hybridized to fragments 73, 33, and 11 was 1:12:14. These results are compatible with a model in which RNA polymerases are stalled just upstream of the C residue at nt 75 (Fig. 1B). In vitro elongation of these stalled RNA polymerases would incorporate one C residue (C-75) into RNA that hybridizes with fragment 73, 13 C residues (C-75 plus 12 C's between nt 72 and 33) into RNA that hybridizes with fragment 33, and 22 C residues (13 C's plus 9 C's between nt 32 and 11) into RNA that hybridizes with fragment 11. Therefore, the ratio of excess counts per minute hybridized to these fragments should be 1:13:22. The value obtained with fragment 11 was not as high as the model predicts, probably because a portion of the stalled RNA polymerases did not elongate in vitro up to the end of that fragment. This is due to the presence of strong in vitro pause sites located between nt 40 and 50, described below (see Fig. 5). In two other experiments, the ratios of excess counts per minute hybridized to these fragments were 1:14:23 and 1:10 (fragment 11 not done). These results are also consistent with the model presented above. If RNA polymerases were stalled just upstream of the pair of C's at nt 81 and 82, the ratio of excess counts per minute in RNA hybridized to fragment 73 and 33 would be 3:15 (1:5), a result that is incompatible with these data. We conclude that the 3′ end of the nascent RNA associated with stalled RNA polymerases is probably located between nt 75 and 80.

TABLE 1. Quantitative analysis of position of stalled RNA polymerases

Fragment	No. of cpm hybridized[a]	No. of C's	No. of cpm/C[b]	No. of excess cpm[c]	Excess cpm/excess for fragment 73
178	448	58	7.72	0	0
119	493	64	7.70	0	0
73	894	80	11.2	276	1
33	3,984	92	43.3	3,274	12
11	4,521	101	44.7	3,740	14

[a] For each sample, three filters containing RNA-DNA hybrids were counted for a total of 15 min in a liquid scintillation counter. Values shown are the average number of counts per minute hybridized minus background radioactivity associated with filters containing M13mp8 DNA.

[b] Number of counts per minute hybridized divided by number of C's.

[c] (Number of counts per minute hybridized to fragment) − (number of counts per minute per C for fragment 178) × (number of C residues in fragment).

Strategy for mapping the distal end of stalled RNA polymerases. Nascent RNAs in transcription complexes can be trimmed by digestion with low concentrations of ribonuclease to within 12 to 25 nt of the growing 3′ end without reduction in the transcriptional activity of the RNA polymerase (30, 41, 43). These 12 to 25 nt are apparently protected from digestion by their association with the RNA polymerase and/or the template DNA strand, although this protection is not absolute: higher concentrations of ribonuclease can remove the nascent RNA up to 3 nt from the growing 3′ end without abolishing transcriptional activity (41).

We set out to map the 5′ ends generated by ribonuclease trimming of nascent RNA associated with stalled RNA polymerases (Fig. 3). After ribonuclease treatment of transcription complexes, RNAs were elongated in the presence of radioactive CTP, and labeled RNA was then hybridized with defined DNA fragments. DNA-RNA hybrids were treated with ribonuclease to remove unpaired RNA tails, and the size of the protected RNA was determined by gel electrophoresis. By this means, the position of the distal end of stalled RNA polymerases can be determined.

To succeed, this approach requires that (i) ribonuclease treatment of the transcription complexes is sufficient to trim nascent RNAs but does not interfere with subsequent elongation by RNA polymerases, (ii) ribonuclease activity is inhibited during the elongation reaction so that the RNAs synthesized remain intact, and (iii) the stalled RNA polymerases are able to extend their nascent RNA chains beyond the end of the DNA fragment used for subsequent hybridization (Fig. 3).

Effect of ribonuclease treatment on the elongation reaction. We found that treatment of viral transcription complexes with 0.1 μg of ribonuclease A per ml for 15 min at 26°C was suffi-

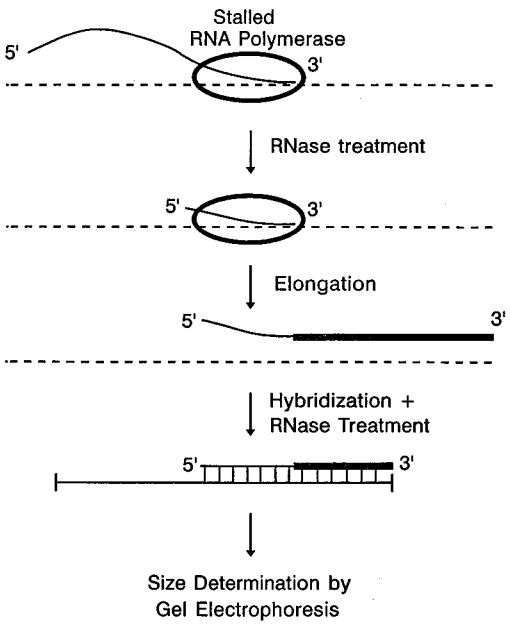

FIG. 3. Strategy used for mapping the distal end of stalled RNA polymerases. The broken line represents the DNA template, the ellipse represents RNA polymerase, and the continuous line represents the nascent RNA chain. Ribonuclease digestion trims the nascent RNA chain. Upon subsequent elongation, the RNA is labeled (thicker line). This RNA is hybridized with a defined single-stranded DNA (bracketed line), and hybrids are digested with ribonuclease to yield a protected RNA of the length defined by the position of its 5′ end.

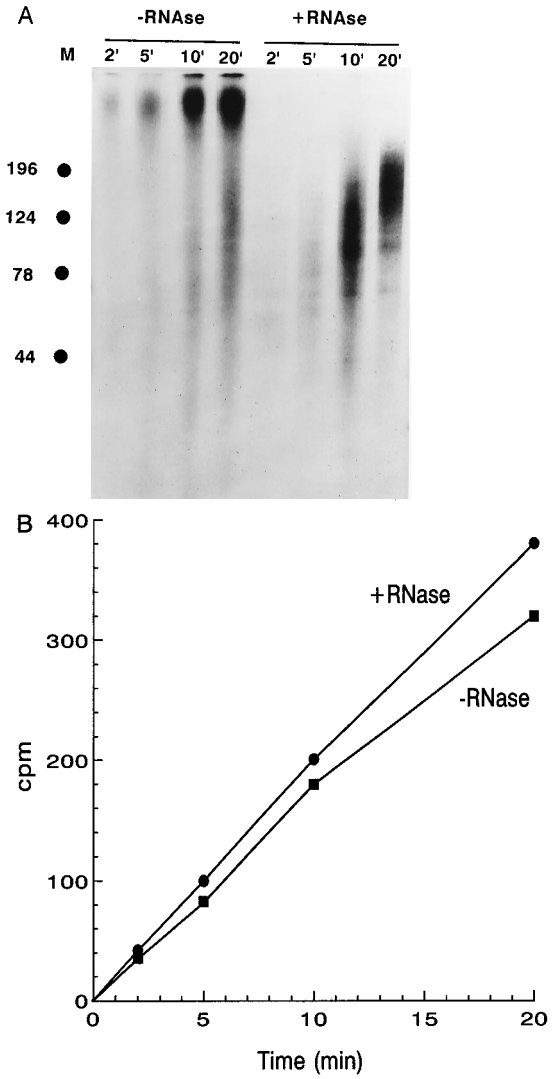

FIG. 4. Effect of ribonuclease treatment on the elongation reaction. (A) Viral transcription complexes were incubated in the absence (−RNAse) or presence (+RNAse) of 0.1 μg of ribonuclease A per ml for 15 min at 26°C. RNA polymerases were then allowed to elongate nascent RNA chains for 2, 5, 10, or 20 min in the presence of 120 U of ribonuclease inhibitor (RNAguard; Pharmacia). RNA was separated from unincorporated nucleotides by exclusion chromatography and analyzed by electrophoresis on an 8% polyacrylamide gel. M, RNA size markers (in nucleotides). (B) Radioactivity in aliquots of each sample was measured in a scintillation counter.

cient to trim nascent RNA tails (Fig. 4A) without significantly affecting the ability of RNA polymerases to elongate nascent RNA in vitro (Fig. 4B). Most of the RNA products generated by untreated viral transcription complexes were at least 400 nt long, independent of the length of the run-on reaction (Fig. 4A, −RNAse). This implies that nascent RNA chains of this length were present before in vitro elongation began. However, there was a minor population of shorter RNAs whose length increased with time of in vitro elongation. Some preparations of viral transcription complexes had very short preexisting RNA chains and/or gave rise to degraded RNA upon in vitro elongation; these preparations were not further used.

After ribonuclease treatment of transcription complexes, elongation (in the presence of a ribonuclease inhibitor) led to production of much shorter RNAs, showing that the ribonuclease treatment was effective (Fig. 4A, +RNAse). These RNAs increased in size by 8 to 10 nt per min at 26°C (with 0.5 μM CTP). There was little sign of degradation of RNA during 20-min run-on reactions. The average size of RNAs elongated for 2 min was 40 to 50 nt, implying that nascent RNAs in ribonuclease-treated transcription complexes were 20 to 30 nt long before elongation began.

Progression of RNA polymerases through the stalling region in vitro. Figure 5 shows the size of RNA chains elongating through the region where stalling occurs during run-ons of from 2 to 20 min with transcription complexes that were not pretreated with ribonuclease. Labeled RNAs were hybridized with a single-stranded DNA fragment spanning nt 173 to 11, the hybrids were treated with ribonuclease to trim unpaired RNA tails, and the protected RNA was analyzed by gel electrophoresis. The autoradiogram shows that RNA polymerases progress through this region in a discontinuous fashion, as revealed by the appearance of RNAs of distinct sizes during elongation. The discontinuous progression of RNA polymerases through a variety of DNA templates has been widely documented (23, 25, 45) and is ascribed to the presence of pause sites at which the progression of RNA polymerases is retarded. The shortest labeled RNA visible in the 2-min run-on was about 100 nt long, which would place its 3′ end near nt 73, assuming that the 5′ ends of these RNAs extended beyond the upstream end of the DNA probe, at nt 173 (an assumption validated by the results shown in Fig. 4A). This would place the 3′ end of nascent RNAs associated with stalled RNA polymerases just upstream of nt 73 before in vitro elongation began, a finding in agreement with the 3′-end analysis described above.

After a 10-min run-on, a band corresponding to the full length of fragment 173-11 appeared (arrow, Fig. 5), indicating that a fraction of the RNA polymerases had progressed up to

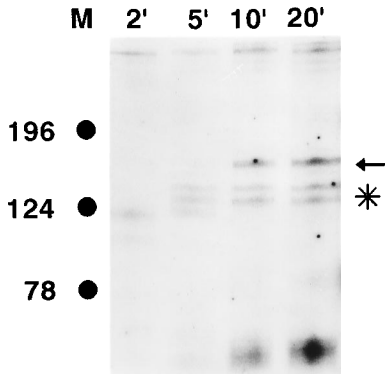

FIG. 5. Progression of RNA polymerases through the stalling region in vitro. In vitro transcription with viral transcription complexes was performed at 26°C for 2, 5, 10, and 20 min, and nascent RNA was hybridized with a DNA fragment spanning nt 173 to 11. RNA protected against ribonuclease A and T_1 digestion by the full-length DNA fragment is indicated by the arrow, and the doublet corresponding to the major in vitro pause sites is indicated by the asterisk. M, RNA size markers (in nucleotides).

or beyond nt 11 by this time. However, a doublet of RNA species 125 to 135 nt long (asterisk, Fig. 5) persisted even after run-ons of 20 min. This suggests that elongating RNA polymerases encounter strong in vitro pause sites around nt 40 and 50. Between nt 54 and 37 of the strand complementary to the late (template) DNA strand, there is a stretch of 18 nt consisting exclusively of C and T residues (Fig. 1, overlining). RNA polymerases may pause here as a result of the limiting concentration of CTP available during in vitro run-ons (45) or because of a tendency of RNA polymerases to pause and/or terminate at runs of T residues in the nontranscribed strand (U residues in the RNA transcript) on natural DNA templates (23–25). Run-ons carried out at 1.3 μM CTP led to an increase in the proportion of RNA polymerases that reached nt 11, but some of the 135-nt RNA remained (data not shown). The presence of these strong pause sites complicates the interpretation of the results presented here, but a significant fraction of RNA polymerases proceeded through these sites to the end of the region used for hybridization. Therefore, the three conditions stated above for carrying out analysis of the trimmed RNAs on ribonuclease-treated transcription complexes were fulfilled.

Mapping the 5′ ends of RNAs generated by ribonuclease treatment of viral transcription complexes. Labeled RNA synthesized by untreated viral transcription complexes during a 10-min run-on was hybridized to DNA fragment 173-11 (Fig. 6A, lane 3). The resulting protected RNAs were full-length read-through (163 nt) and paused species (135 nt) as previously shown in Fig. 5. In this experiment, the 135-nt species (paused at nt 40) was predominant; little of the 125-nt species (paused at nt 50) was seen. The corresponding RNA from ribonuclease-treated transcription complexes (Fig. 6A, lane 4) gave rise to faster-migrating species approximately 100 and 70 nt long. The 5′ ends of these RNAs therefore lie near nt 110, some 60 nt downstream from the end of fragment 173-11 (Fig. 6B).

To substantiate this conclusion, the same RNAs were hybridized with DNA fragments 406-33 and 406-11. RNAs from untreated transcription complexes were protected over the full length of these DNA fragments, as expected (Fig. 6A, lanes 1 and 2). RNAs from ribonuclease-treated transcription complexes gave rise to the same 100- and 70-nt protected species when hybridized with fragment 406-11 (lane 5) as when hybridized with fragment 173-11 (lane 4). However, hybridization

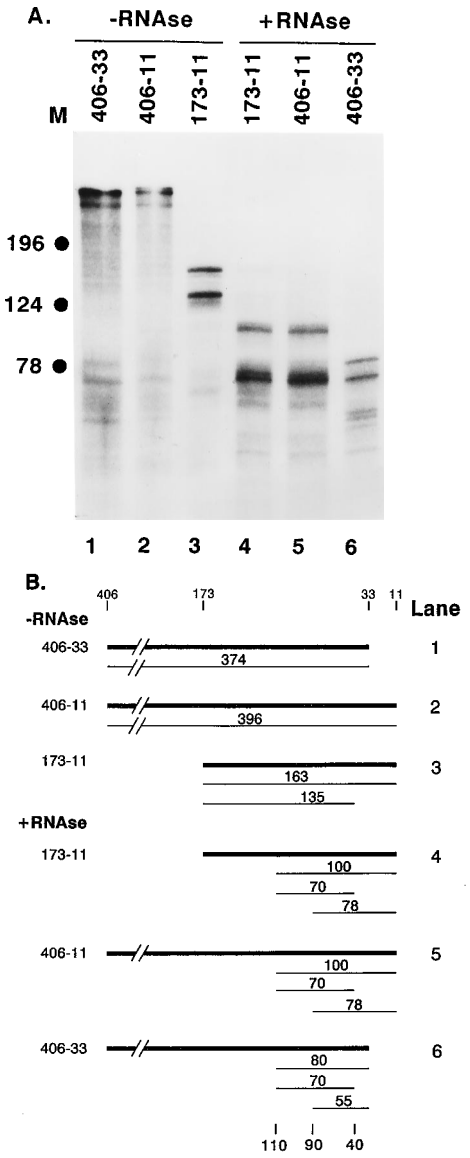

FIG. 6. Mapping nascent RNA protected against ribonuclease digestion by stalled RNA polymerases. (A) Viral transcription complexes were incubated without (−RNAse [lanes 1 to 3]) or with (+RNAse [lanes 4 to 6]) 0.1 μg of ribonuclease A per ml prior to a 10-min run-on. RNA products were hybridized with DNA fragment 173-11 (lanes 3 and 4), 406-11 (lanes 2 and 5), or 406-33 (lanes 1 and 6). (B) Schematic representation of the DNA fragments used (thick lines) and of the RNAs identified (thin lines). The lengths of each RNA, estimated from migration compared with marker RNAs and/or from the lengths of the corresponding DNA fragments, are noted above each line. The numbers at the top of the figure show the ends of the DNA fragments; the numbers at the bottom of the figure show the positions of the 5′ and 3′ ends of RNAs as described in text.

with fragment 406-33 led to a reduction in the longest protected RNA from 100 nt to about 80 nt, as would be expected if this RNA had a 5′ end at nt 110 (Fig. 6B). RNAs produced by polymerases that pause in vitro between nt 50 and 40 would not be expected to change in size when hybridized with these three DNA fragments. This prediction is confirmed by the presence of the 70-nt species in lanes 4, 5, and 6.

On the other hand, a protected RNA species about 78 nt long, detected in lanes 4 and 5 as a band migrating just above the 70-nt species, was not present after hybridization with

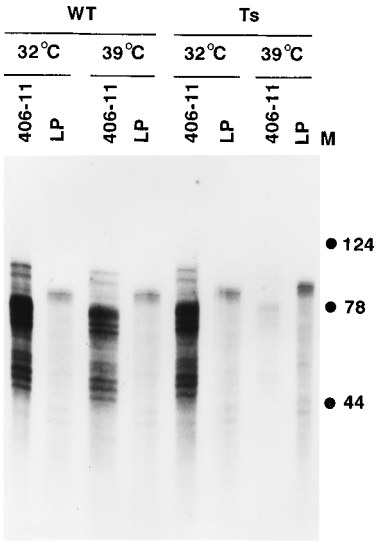

FIG. 7. Mapping of stalled RNA polymerases with a thermosensitive virus strain. Viral transcription complexes were isolated from cells infected with either the wild-type AT3-Modori strain (WT) or a thermosensitive strain, AT3ts25E (Ts), incubated for 72 h at the permissive temperature (32°C), and then shifted to the nonpermissive temperature (39°C) for 2.5 h. Viral transcription complexes were incubated with 0.1 μg of ribonuclease A per ml prior to a 10-min run-on. RNA products were hybridized with the DNA fragment 406-11 or with a DNA fragment (LP) spanning the late promoter region (nt 5134 to 4968). M, RNA size markers (in nucleotides).

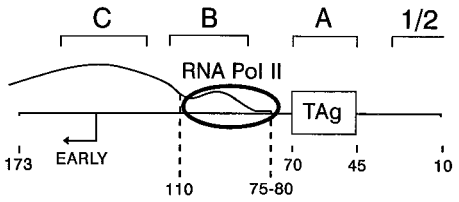

FIG. 8. Location of RNA polymerases stalled upstream of binding site A. RNA polymerase II (Pol II [ellipse]) is shown protecting nascent RNA (curved line) between nt 75 to 80 and 110, just upstream of large T antigen (TAg) bound to site A on viral DNA. The locations of other large T antigen binding sites, as well as the site for initiation of early transcription, are shown.

fragment 406-33 (lane 6); several faster-migrating species 50 to 55 nt long appeared in its place. The origin of these species is not clear, but they may arise from a second population of stalled RNA polymerases located some 20 to 25 nt downstream of the major population. The distal end of these polymerases would lie near nt 90, as shown in Fig. 6B.

A fraction of the nascent RNA associated with stalled RNA polymerases was already trimmed before ribonuclease treatment, as shown by the presence of small amounts of 100-, 80-, and 70-nt RNA species made by untreated transcription complexes (Fig. 6A, lanes 1 to 3). These RNAs could be generated either by digestion by endogenous ribonucleases in the cell or by degradation during isolation of viral transcription complexes.

The 100-nt RNA is generated by stalled RNA polymerases. To confirm that the RNAs we have analyzed originate from stalled RNA polymerases, we examined the temperature dependence of their synthesis by using a temperature-sensitive virus, AT3ts25E (7). This strain has a substitution of cysteine for glycine in the carboxy terminus of the large T antigen (16, 47). This mutation renders the protein thermolabile at the nonpermissive temperature (39°C). Bertin et al. (7) showed that excess RNA polymerases accumulated in the intergenic region of the AT3ts25E strain at the permissive temperature, while no accumulation occurred when the cells were shifted to the nonpermissive temperature for 2 to 5 h.

Cells were infected with either the wild-type strain or AT3ts25E virus and incubated for 72 h at 32°C. The cells were then maintained at this temperature or incubated at 39°C for an additional 2.5 h. Viral transcription complexes were then isolated, treated with ribonuclease A, and allowed to elongate nascent RNA chains for 10 min. The RNA products were hybridized with DNA fragment 406-11 (Fig. 7). As expected, there were no differences between the patterns seen at 32 and 39°C with the wild-type virus. The 100-nt band and the bands corresponding to the in vitro pause site were seen. The ratio between full-length RNA and in vitro paused RNA varies between preparations of viral transcription complexes. In these experiments, less full-length RNA was made. Infection with the thermosensitive virus at 32°C led to the same pattern as that with wild-type virus. However, when the temperature was shifted to 39°C, at which temperature large T antigen is no longer functional, RNAs hybridizing with fragment 406-11 were strongly reduced. This shows that the RNAs we have analyzed arise from stalled RNA polymerases, because they depend on the presence of functional large T antigen (7).

It could be argued that loss of the signal at 39°C with the AT3ts25E virus is not due to loss of RNA polymerase stalling but instead is due to an overall decrease in late promoter activity. To check this, we hybridized run-on RNA with a DNA fragment overlapping the late promoter. An RNA 80 nt long was protected by this fragment, and no significant difference in its amount was detected with either wild-type or temperature-sensitive virus after shift-up. Thus, the loss of the signal with the AT3ts25E virus at 39°C was not caused by a decrease in the level of late promoter activity but indeed was a result of the inability of inactivated large T antigen to block progression of RNA polymerase II through the origin region in vivo.

DISCUSSION

Polyomavirus large T antigen is capable of blocking elongation by RNA polymerase II in vivo when bound to site A in the intergenic region of polyomavirus DNA. In this report, we present an analysis of the ternary elongation complex involving RNA polymerase II stalled in vivo near site A. We mapped the location of the stalled RNA polymerases with a precision of 5 nt or less by analyzing the nascent RNA chains associated with these RNA polymerases. The stalled polymerases lie immediately upstream of the DNAse I footprint of large T antigen binding to site A (Fig. 8). The 3′ end of the nascent RNA is 5 to 10 nt upstream of site A, and the 5′ end of the RNA protected against ribonuclease A digestion is 40 nt upstream of site A. The stalled RNA polymerases therefore protect an RNA fragment about 30 nt long against ribonuclease A digestion. The proximity of the 3′ end of the nascent RNA to the footprint at site A suggests a very close interaction between the front end of stalled RNA polymerase II and DNA-bound polyomavirus large T antigen.

Until recently, technical difficulties have prevented detailed examination of the structure of the elongating complex composed of RNA polymerase II, the DNA template, and the nascent RNA transcript. A crucial component of our understanding of the mechanism of transcriptional elongation and termination is the determination of the structure and properties of this ternary complex. It has now become evident that the structure of RNA polymerases moving along a transcription

unit is not static but instead undergoes constant changes. Work by Krummel and Chamberlin (27, 28) with *E. coli* RNA polymerase showed that DNAse I footprints of ternary complexes, paused artificially by deprivation of specific nucleotides, were of different sizes, even though these complexes differed from one another by only one or a few nucleotides on the nascent RNA. These observations led to the "inchworm model" of transcription elongation, in which the distal (rear) edge of the footprint moves more or less simultaneously with the progression of the 3′ end of the RNA, while the proximal (front) edge moves by steps, leading to compression and expansion of the footprint during transcription. Only a few cases of ternary complexes involving eukaryotic RNA polymerases have been studied so far. DNAse I footprint studies suggest that RNA polymerase II also undergoes structural changes during the course of transcription (32, 40).

For *E. coli* RNA polymerase, the distance between the proximal edge of the footprint and the 3′ end of the nascent RNA has been shown to be relatively constant at 7 to 9 nt when complexes are paused over similar sequences (35). This distance is more variable (6 to 14 nt) when complexes are halted over different sequences (27, 28, 37). Thus, the movement of the active site within the RNA polymerase molecule seems to depend on the sequences being transcribed. Pavco and Steege (38) showed that SP6 and T7 RNA polymerases were able to extend the 3′ end of the nascent RNA up to 5 to 7 nt into the footprint of the DNA-binding protein (*Eco*RI) used to block progression of these polymerases. Since these polymerases are formed by a single polypeptide (11, 36), instead of a complex of multiple polypeptides as for the *E. coli* or eukaryotic RNA polymerases (52), they might be able to move further into the DNA-bound protein because of their small size. The distance between the 3′ end of the RNA and the front end of RNA polymerase II also varies among different paused ternary complexes (32, 40). This variation suggests that the catalytic site for RNA synthesis (i.e., the site in contact with the 3′ end of the nascent RNA) is flexible, since its position relative to the mass of the polymerase is variable. Our results show that under conditions in which RNA polymerase II has stalled naturally in vivo, the 3′ end of the nascent RNA can come as close as 5 to 10 nt from the footprint of the blocking protein.

Under the conditions we have used, the nascent RNA was protected by RNA polymerase II against ribonuclease A digestion over approximately 30 nt. This is consistent with results from a previous report in which RNA polymerase II progression was blocked in vitro by the lack of a specific nucleotide and reinitiation was allowed to occur (44). In that study, several RNA polymerases accumulated upstream of the blockage site on each template DNA molecule, and the distance between each successive RNA 3′ end was about 30 nt. Incidentally, we saw no evidence for the accumulation of more than one RNA polymerase II molecule at the stalling site in the present study. In another report, elongation by *E. coli* RNA polymerase was stopped by being chilled at 0°C and addition of 50 mM EDTA; a length of 22 ± 4 nt of the nascent RNA chain was protected against ribonuclease A digestion (43).

The presence of an RNA-DNA hybrid of 12 bp in the ternary complex has been proposed in several reports (5, 18, 22, 30). However, this claim has been challenged by two lines of evidence. Rice et al. (41) studied the size of RNA molecules protected against digestion by high concentrations of ribonucleases A and T_1 by using artificially paused calf thymus RNA polymerase II in vitro. Their results suggested the presence of a very short RNA-DNA hybrid no more than 2 or 3 nt long. Lee and Landick (31) mapped the region of template DNA melted by paused *E. coli* RNA polymerase, with a $KMnO_4$ base modification technique. They showed that an 8-bp region on the template strand was protected against $KMnO_4$ modification and attributed this protection to the presence of an RNA-DNA hybrid. Whatever the real size of the RNA-DNA hybrid in elongating transcription complexes is, it is unlikely that the entire 30 nt protected in our studies is present as an RNA-DNA hybrid. Instead, a portion of the nascent RNA is likely to be protected via binding to the RNA polymerase molecule. Such an interaction is now well established; RNA polymerase II is in fact capable of interacting with an RNA transcript in vitro in the absence of DNA (10, 21).

Data shown in Fig. 6 suggest the presence of a second population of stalled RNA polymerases. These polymerases might accumulate as a result of encountering large T antigen bound only to the downstream copy of the two GAGGC target binding sequences in site A (shown as the complementary CGGTC sequences underlined by arrows in Fig. 1B) or perhaps large T antigen bound to site 1/2 farther downstream. Further work, with viruses with mutations in one or both GAGGC sequences in site A, will be necessary to distinguish between these alternatives. In our previous work (6), the involvement of site 1/2 in RNA polymerase stalling was not examined, since integrity of this binding site is essential for viral DNA replication and therefore for virus viability. Large T antigen binds as a complex double-hexamer structure to this site, triggering DNA unwinding (9, 51) and leading to initiation of DNA replication. RNA polymerases transcribing through this region could conceivably dislodge DNA-binding proteins necessary for initiation of DNA replication, in a process analogous to the inactivation of transcriptional promoters by transcription from upstream sites (promoter occlusion [14, 19, 34, 49]). It is therefore possible that the function of the block to RNA polymerase elongation at site A is to protect the polyomavirus origin of replication against read-through transcription.

The data presented in this paper represent the first structural analysis of RNA polymerase II ternary complexes stalled in vivo at a site known to be occupied by a DNA-binding protein. Since all regulatory factors involved in transcription were presumably present when these complexes stalled within the intact cell, they are likely to be a faithful representation of the elongation complex present in vivo. The strategies we used here targeted ternary complexes capable of resuming RNA chain synthesis in vitro, therefore allowing us to analyze specifically the active ternary complexes representative of elongation complexes in vivo.

ACKNOWLEDGMENTS

We thank Lynn Kelly, Sandra Iacampo, and Hongling Li for excellent technical assistance.

This research was supported by grant MT-7281 from the Medical Research Council of Canada. F.B. was supported by an F. C. Harrison Fellowship, studentships from le Fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec (FCAR), and the Medical Research Council of Canada.

REFERENCES

1. **Acheson, N. H.** 1978. Polyoma virus giant RNAs contain tandem repeats of the nucleotide sequence of the entire genome. Proc. Natl. Acad. Sci. USA **75:**4754–4758.
2. **Acheson, N. H.** 1984. Kinetics and efficiency of polyadenylation of late polyomavirus nuclear RNA: generation of oligomeric polyadenylated RNAs and their processing into mRNA. Mol. Cell. Biol. **4:**722–729.
3. **Acheson, N. H., E. Buetti, K. Scherrer, and R. Weil.** 1971. Transcription of the polyoma virus genome: synthesis and cleavage of giant late polyoma-specific RNA. Proc. Natl. Acad. Sci. USA **68:**2231–2235.
4. **Ashfield, R., P. Enriquez-Harris, and N. J. Proudfoot.** 1991. Transcriptional termination between the closely linked human complement genes C2 and

factor B: common termination factor for C2 and *c-myc*? EMBO J. **10:**4197–4207.
5. **Bernhard, S. L., and C. F. Meares.** 1986. Accessibility of the leading end of ribonucleic acid in transcription complexes. Biochemistry **25:**6397–6404.
6. **Bertin, J., N.-A. Sunstrom, and N. H. Acheson.** 1993. Mutation of large T-antigen-binding site A, but not site B or C, eliminates stalling by RNA polymerase II in the intergenic region of polyomavirus DNA. J. Virol. **67:**5766–5775.
7. **Bertin, J., N.-A. Sunstrom, P. Jain, and N. H. Acheson.** 1992. Stalling by RNA polymerase II in the polyomavirus intergenic region is dependent on functional large T antigen. Virology **189:**715–724.
8. **Birg, F., J. Favaloro, and R. Kamen.** 1977. Analysis of polyoma virus nuclear RNA by mini-blot hybridization. Proc. Natl. Acad. Sci. USA **74:**3138–3142.
9. **Borowiec, J. A., F. B. Dean, P. A. Bullock, and J. Hurwitz.** 1990. Binding and unwinding—how T antigen engages the SV40 origin of DNA replication. Cell **60:**181–184.

9a. **Brabant, F., and N. H. Acheson.** Unpublished data.

10. **Busby, S., A. Spassky, and H. Buc.** 1981. On the binding of tRNA to *Escherichia coli* RNA polymerase. Interactions between the core enzyme, DNA and tRNA. Eur. J. Biochem. **118:**443–451.
11. **Butler, E. T., and M. J. Chamberlin.** 1982. Bacteriophage SP6-specific RNA polymerase. I. Isolation and characterization of the enzyme. J. Biol. Chem. **257:**5772–5778.
12. **Connelly, S., and J. L. Manley.** 1989. A CCAAT box sequence in the adenovirus major late promoter functions as part of an RNA polymerase II termination signal. Cell **57:**561–571.
13. **Connelly, S., and J. L. Manley.** 1989. RNA polymerase II transcription termination is mediated specifically by protein binding to a CCAAT box sequence. Mol. Cell. Biol. **9:**5254–5259.
14. **Corbin, V., and T. Maniatis.** 1989. Role of transcriptional interference in the *Drosophila melanogaster Adh* promoter switch. Nature (London) **337:**279–282.
15. **Cowie, A., and R. Kamen.** 1984. Multiple binding sites for polyomavirus large T antigen within regulatory sequences of polyomavirus DNA. J. Virol. **52:**750–760.
16. **Deininger, P. L., P. LaPorte, and T. Friedmann.** 1981. Nucleotide sequence changes in polyoma *ts-a* mutants: correlation with protein structure. J. Virol. **37:**871–875.
17. **Dilworth, S. M., A. Cowie, R. I. Kamen, and B. E. Griffin.** 1984. DNA binding activity of polyoma virus large tumor antigen. Proc. Natl. Acad. Sci. USA **81:**1941–1945.
18. **Hanna, M. H., and C. F. Meares.** 1983. Topography of transcription: path of the leading end of nascent RNA through the *Escherichia coli* transcription complex. Proc. Natl. Acad. Sci. USA **80:**4238–4242.
19. **Henderson, S. L., K. Ryan, and B. Sollner-Webb.** 1989. The promoter-proximal rDNA terminator augments initiation by preventing disruption of the stable transcription complex caused by polymerase read-in. Genes Dev. **3:**212–223.
20. **Hyde-deRuyscher, R. P., and G. G. Carmichael.** 1990. Polyomavirus late pre-mRNA processing: DNA replication-associated changes in leader exon multiplicity suggest a role for leader-to-leader splicing in the early-late switch. J. Virol. **64:**5823–5832.
21. **Johnson, T. L., and M. J. Chamberlin.** 1994. Complexes of yeast RNA polymerase II and RNA are substrates for TFIIS-induced RNA cleavage. Cell **77:**217–224.
22. **Kainz, M., and J. Roberts.** 1992. Structure of transcription elongation complexes *in vivo*. Science **255:**838–841.
23. **Kerppola, T. K., and C. M. Kane.** 1988. Intrinsic sites of transcription termination and pausing in the c-*myc* gene. Mol. Cell. Biol. **8:**4389–4394.
24. **Kerppola, T. K., and C. M. Kane.** 1990. Analysis of the signals for transcription termination by purified RNA polymerase II. Biochemistry **29:**269–278.
25. **Kessler, M., E. Ben-Asher, and Y. Aloni.** 1989. Elements modulating the block of transcription elongation at the adenovirus 2 attenuation site. J. Biol. Chem. **264:**9785–9790.
26. **Koenig, M., E. P. Hoffman, C. J. Bertelson, A. P. Monaco, C. Feener, and L. K. Kunkel.** 1987. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell **50:**509–517.
27. **Krummel, B., and M. J. Chamberlin.** 1992. Structural analysis of ternary complexes of *Escherichia coli* RNA polymerase. Individual complexes halted along different transcription units have distinct and unexpected biochemical properties. J. Mol. Biol. **225:**221–237.
28. **Krummel, B., and M. J. Chamberlin.** 1992. Structural analysis of ternary complexes of *Escherichia coli* RNA polymerase. Deoxyribonuclease I footprinting of defined complexes. J. Mol. Biol. **225:**239–250.
29. **Kuhn, A., I. Bartsch, and I. Grummt.** 1990. Specific interaction of the murine transcription termination factor TTF I with class-I RNA polymerases. Nature (London) **344:**559–562.
30. **Kumar, S. A., and J. S. Krakow.** 1975. Studies on the product binding site of the *Azotobacter vinelandii* ribonucleic acid polymerase. J. Biol. Chem. **250:**2878–2884.
31. **Lee, D. N., and R. Landick.** 1992. Structure of RNA and DNA chains in paused transcription complexes containing *Escherichia coli* RNA polymerase. J. Mol. Biol. **228:**759–777.
32. **Linn, S. C., and D. S. Luse.** 1991. RNA polymerase II elongation complexes paused after the synthesis of 15- or 35-base transcripts have different structures. Mol. Cell. Biol. **11:**1508–1522.
33. **Llopis, R., F. Perrin, F. Bellard, and P. Gariglio.** 1981. Quantitation of transcribing native simian virus 40 minichromosomes extracted from CV1 cells late in infection. J. Virol. **38:**82–90.
34. **McStay, B., and R. H. Reeder.** 1990. An RNA polymerase I termination site can stimulate the adjacent ribosomal gene promoter by two distinct mechanisms in *Xenopus laevis*. Genes Dev. **4:**1240–1252.
35. **Metzger, W., P. Schickor, and H. Heumann.** 1989. A cinematographic view of *Escherichia coli* RNA polymerase translocation. EMBO J. **8:**2745–2754.
36. **Moffatt, B. A., J. J. Dunn, and F. W. Studier.** 1984. Nucleotide sequence of the gene for bacteriophage T7 RNA polymerase. J. Mol. Biol. **173:**265–269.
37. **Pavco, P. A., and D. A. Steege.** 1990. Elongation by *Escherichia coli* RNA polymerase is blocked *in vitro* by a site-specific DNA binding protein. J. Biol. Chem. **265:**9960–9969.
38. **Pavco, P. A., and D. A. Steege.** 1991. Characterization of elongating T7 and SP6 RNA polymerases and their response to a roadblock generated by a site-specific DNA binding protein. Nucleic Acids Res. **19:**4639–4646.
39. **Pomerantz, B. J., C. R. Mueller, and J. A. Hassell.** 1983. Polyomavirus large T antigen binds independently to multiple, unique regions on the viral genome. J. Virol. **47:**600–610.
40. **Rice, G. A., M. J. Chamberlin, and C. M. Kane.** 1993. Contacts between mammalian RNA polymerase II and the template DNA in a ternary elongation complex. Nucleic Acids Res. **21:**113–118.
41. **Rice, G. A., C. M. Kane, and M. J. Chamberlin.** 1991. Footprinting analysis of mammalian RNA polymerase II along its transcript: an alternative view of transcription elongation. Proc. Natl. Acad. Sci. USA **88:**4245–4249.
42. **Roberts, S., T. Purton, and D. L. Bentley.** 1992. A protein-binding site in the *c-myc* promoter functions as a terminator of RNA polymerase II transcription. Genes Dev. **6:**1562–1574.
43. **Rohrer, H., and W. Zillig.** 1977. Studies on the transcription complex of *Escherichia coli* RNA polymerase. Eur. J. Biochem. **79:**401–409.
44. **Sawadogo, M., and R. G. Roeder.** 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43:**165–175.
45. **Skarnes, W. C., and N. H. Acheson.** 1991. RNA polymerase II pauses *in vitro*, but does not terminate, at discrete sites in promoter-proximal regions on polyomavirus transcription complexes. Virology **182:**54–60.
46. **Skarnes, W. C., D. C. Tessier, and N. H. Acheson.** 1988. RNA polymerases stall and/or prematurely terminate nearby both early and late promoters on polyomavirus DNA. J. Mol. Biol. **203:**153–171.
47. **Thomas, T., P. Vollmer, and W. R. Folk.** 1981. Nucleotide sequence changes in polyoma virus *A* gene mutants. J. Virol. **37:**1094–1098.
48. **Tseng, R. W., and N. H. Acheson.** 1986. Use of a novel S1 nuclease RNA-mapping technique to measure efficiency of transcription termination on polyomavirus DNA. Mol. Cell. Biol. **6:**1624–1632.
49. **Vales, L. D., and J. E. Darnell, Jr.** 1989. Promoter occlusion prevents transcription of adenovirus polypeptide IX mRNA until after DNA replication. Genes Dev. **3:**49–59.
50. **van Holde, K. E., D. E. Lohr, and C. Robert.** 1992. What happens to nucleosomes during transcription? J. Biol. Chem. **267:**2837–2840.
51. **Wang, E. H., and C. Prives.** 1991. ATP induces the assembly of polyoma large T antigen into hexamers. Virology **184:**399–403.
52. **Young, R. A.** 1991. RNA polymerase II. Annu. Rev. Biochem. **60:**689–715.