

Oligonucleotide sequence signaling transcriptional termination of vaccinia virus early genes

(RNA polymerase/*in vitro* transcription/poxvirus)

LEONARD YUEN AND BERNARD MOSS*

Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

Communicated by Donald D. Brown, June 15, 1987

ABSTRACT In an *in vitro* system containing enzymes extracted from vaccinia virions, transcription of the vaccinia growth factor gene terminated ≈50 base pairs downstream of a thymidine-rich sequence. Deletion mutagenesis suggested the presence of two tandem termination signals. The signal was identified by replacing the 3' end of the gene with the oligonucleotide AATTTTAT that induced downstream termination. Further analysis of the transcripts formed with a series of templates containing 16 related synthetic oligonucleotides established the minimum functional termination signal as TTTTNT, in which N represents any nucleotide. Termination efficiency may be increased, however, by the presence of an adenine preceding the thymidine cluster. The general use of this signal at early times in infection but not at late times is supported by a survey of vaccinia virus gene sequences.

Of the various steps in eukaryotic mRNA biogenesis—initiation, elongation, termination, 3' processing, capping, polyadenylation, and splicing—probably least is known about termination. The available information suggests that termination is heterogeneous and may occur far downstream of the mature, processed 3' end of the transcript (1–3). Several putative termination sequences have been proposed (4–8), but in general processing and termination mechanisms have not been adequately distinguished because of difficulties in developing an *in vitro* system. However, termination was shown to occur in an *in vitro* system derived from vaccinia virus particles (9).

Vaccinia virus is a large DNA virus that replicates in the cytoplasm, synthesizes capped and polyadenylated mRNAs, encodes its own multisubunit RNA polymerase with sequence homology to cellular RNA polymerases, and packages the entire transcription system in infectious virus particles (10, 11). The latter feature considerably facilitated the preparation of an *in vitro* transcription system. Both the 5' and 3' ends of transcripts made *in vitro* (9, 12) corresponded to those made *in vivo* (13). Termination was distinguished from 3' processing by kinetic analysis, by characterization of RNA products, and by the absence of cleavage of extended transcripts that were added to the transcription system. The termination signal of the vaccinia virus growth factor (VGF) gene was localized to a region containing runs of thymidine residues ≈50 base pairs (bp) upstream of the site of termination. We now show, by deletion mutagenesis of the natural termination sequence and by replacing it with synthetic oligonucleotides, that the minimum signal for termination is TTTTNT, where the N is any nucleotide.

MATERIALS AND METHODS

In Vitro Transcription Reactions. Enzymes used for *in vitro* transcription were prepared by deoxycholate extraction of

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

vaccinia virus particles, DEAE-cellulose chromatography to remove DNA, and glycerol gradient centrifugation (14). Standard 50-μl transcription reactions contained 20 μM Hepes (pH 8.0), 1 mM dithiothreitol, 4 mM MgCl₂, 1 mM ATP, 1 mM UTP, 1 mM CTP, 0.1 mM GTP, 5 μCi [α-³²P]GTP (800 Ci/mmol; 1 Ci = 37 GBq; Amersham), 5 μl of glycerol gradient-isolated enzyme, and 100 ng of DNA template. Transcription was carried out at 30°C for 1 hr, and the RNA was extracted and precipitated as described (9).

Analysis of Transcription Products. The ethanol precipitated RNA was dissolved in 80% (vol/vol) formamide, heated at 70°C, and electrophoresed on a 4% polyacrylamide/8 M urea minigel as described (9).

Construction of Deletion Mutants. Plasmid pLY1 containing the entire VGF gene was used to construct 3'→5' deletion mutants (9). To prepare 5'→3' deletions, pLY1 was linearized by cleavage with Acc I, which cuts in the middle of the VGF gene (Fig. 1). The DNA was then incubated successively with exonuclease III, S1 nuclease, and Klenow fragment of DNA polymerase (4). The DNA was then digested with HindIII, which cuts in vector DNA upstream of the VGF gene. The exonuclease digested pLY1 large DNA fragment containing the 3' half of the VGF gene, and most of the vector DNA was purified by agarose gel electrophoresis. The small fragment was purified from pLY1 that had been treated successively with Acc I, Klenow fragment of DNA polymerase, and HindIII without undergoing exonuclease digestion. By ligating the undigested small fragment and the exonuclease digested large fragment, a set of circular plasmids containing unidirectional 5'→3' deletions from the Acc I site was obtained. The plasmids were used to transform HB101 cells, and ampicillin-resistant transformants were isolated. The extent of deletion was first estimated by restriction endonuclease analysis and by agarose gel electrophoresis and then determined precisely by sequencing.

Construction of Templates Containing Synthetic Oligonucleotides. Plasmid pLY1 was digested with Acc I and EcoRI to remove the 3'-terminal portion of the VGF gene. Synthetic complementary oligonucleotides with Acc I and EcoRI ends were made in an Applied Biosystems 380B DNA synthesizer and ligated to the large Acc I-EcoRI fragment of pLY1. The oligonucleotides were not phosphorylated to ensure single-copy insertion into the vector. Ampicillin-resistant transformants were isolated, and the DNA was analyzed by restriction enzyme digestion and DNA sequencing.

RESULTS

Effects of Deletions on Transcription Termination. Transcription termination was demonstrated using a crude extract of vaccinia virus as the source of enzymes and a linear DNA fragment containing the entire VGF gene as template (9). For

Abbreviations: VGF, vaccinia virus growth factor; nt, nucleotide.

*To whom reprint requests should be addressed at: Building 5, Room 324, National Institutes of Health, Bethesda, MD 20892.

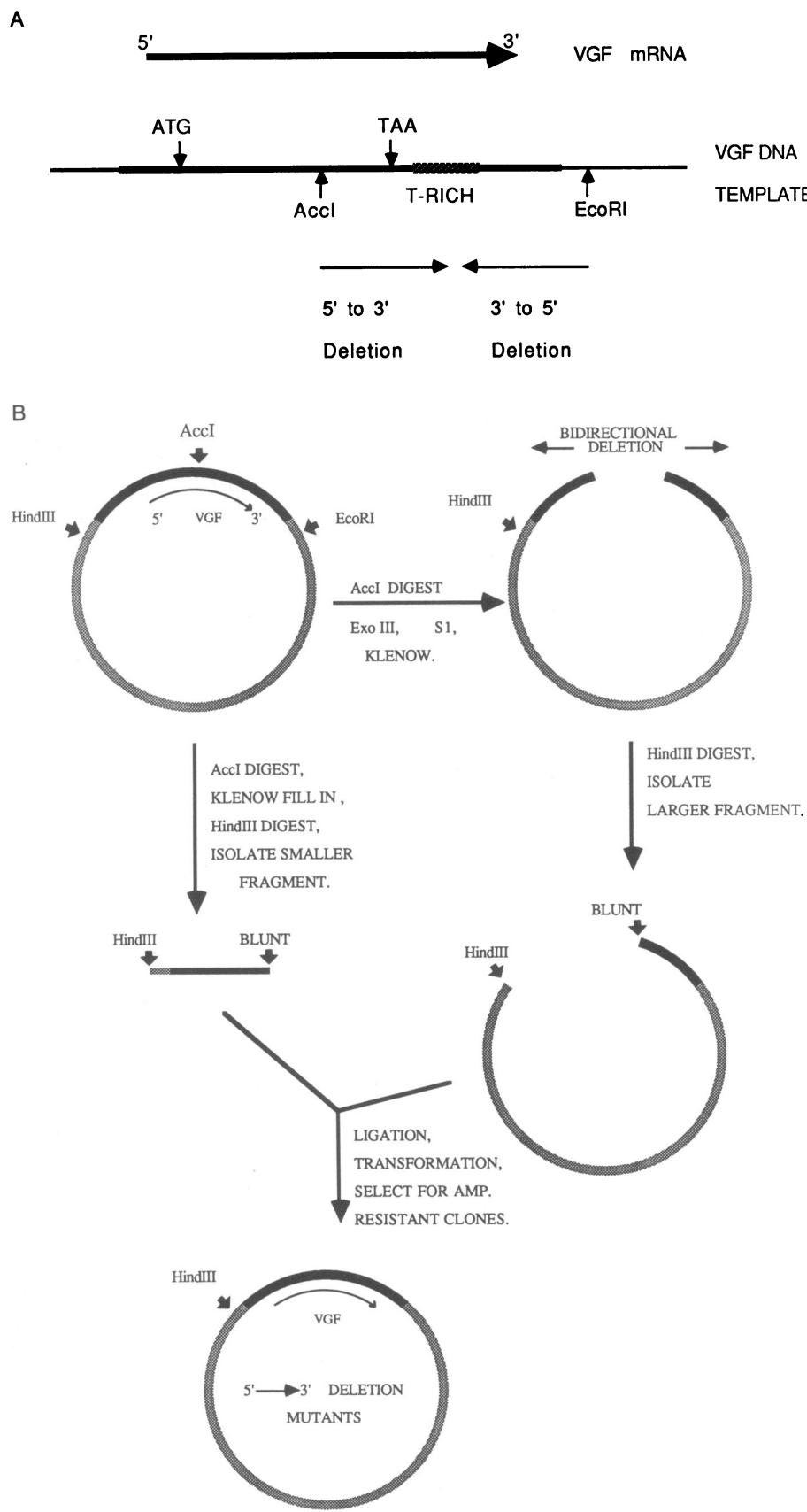


FIG. 1. Construction of 5'→3' deletion mutants. (A) The VGF template and the expected mRNA transcript are shown. The viral sequence (thick line) containing the translation initiation (ATG) and termination (TAA) codons and the thymidine-rich sequence in the 3'-noncoding region of the gene are shown. The pUC plasmid sequence surrounding the viral sequence is represented by the thin line. Sites (Acc I and EcoRI) at which deletions start are indicated. (B) The strategy for constructing 5'→3' deletion mutants is shown schematically. AMP, ampicillin.

the present study, transcription was carried out with glycerol gradient fractions containing RNA polymerase and associated factors that were extracted from vaccinia virus particles (14), and the VGF gene was again used as template (Fig. 1A). This system initiates and terminates transcription as well as

the crude system but is largely free of poly(A) polymerase, which can cause elongation and size variability of RNA. The terminated and read-through VGF transcripts are 530 and 820 nucleotides (nt), respectively; the third 250-nt band is a run-off product of an oppositely oriented nonoverlapping

second gene present in truncated form on the template (Fig. 2A, lane 5). The efficiency of termination was calculated by densitometry of the autoradiographs of [α -³²P]GTP-labeled RNAs and corrected for the greater length and higher guanosine composition of the read-through compared to the terminated RNA species. Termination efficiencies of 60–70% were generally obtained.

By preparing a set of templates with 3'→5' deletions, at least part of the termination signal for the VGF gene was shown to be included within a thymidine-rich region between the actual site of transcription termination and the translation stop codon (9). This region, shown in Figs. 1A and 2, contains three clusters of five, seven, and five thymidines. The TAA triplet near the left end of the indicated sequence is the translation stop codon of the VGF gene. Termination continued to occur resulting in a 530-nt transcript even after both the second and third clusters of thymidines were removed (Fig. 2A, lanes 3 and 4). Differences in the overall rates of RNA synthesis, including that of the 250-nt species, resulted from difficulties in accurately estimating the concentrations of purified DNA fragments used as templates. This problem was minimized by comparing the ratios rather than the absolute amounts of terminated and read-through VGF transcripts. The efficiency of termination, relative to the original template, was 75% when only the first block of thymidines remained. The lack of termination when all three blocks were removed (Fig. 2A, lanes 1 and 2) is consistent with data obtained using a crude extract of vaccinia virus (9).

Although the above results demonstrate the importance of the first thymidine cluster, they do not indicate whether additional upstream sequences are necessary for termination. Therefore, another set of templates was prepared with deletions starting at an *Acc* I site, ≈240 bp upstream of the 3' end of the VGF gene, and continuing for various distances into the thymidine-rich region (Fig. 1B). With these templates, the terminated RNA should be ≈380 nt long and the read-through ≈640 nt long. Deletions up to the stop codon had no significant effect on termination indicating that the signal did not start within the coding region. Paradoxically, deletion of the first run of thymidines had little or no effect

on termination (Fig. 2B, lane 3), whereas deletion of both the first and second blocks resulted in formation of read-through transcripts only (Fig. 2B, lanes 4 and 5). With the first block of thymidines removed, the efficiency of termination, relative to the original template, was 85%. One interpretation of these data is that sequences associated with either of the first two sets of thymidines can serve as independent termination signals but that the third cannot.

Construction of the Putative Termination Sequence with Synthetic Oligonucleotides. To eliminate entirely the possible role of additional sequences adjacent to the thymidine-rich region, a 300-bp fragment containing the 3' portion of the VGF gene was deleted by cleavage with *Acc* I (which cuts 240 bp upstream of the 3' end of the VGF gene) and *Eco*RI (which cuts immediately downstream of the VGF gene in the vector DNA) and replaced with a double-stranded oligonucleotide corresponding to the first cluster AATTCTTAT (with added *Acc* I and *Eco*RI overhanging ends). With this template, the efficiency of termination was 94% of that obtained with the original one (Fig. 3A, lane 2). Thus, no other specific sequences near the 3' end of the VGF gene are required for termination. It is important to point out that the actual site of termination is always downstream of the signal and was in plasmid DNA with this template as well as with the templates in which 3'→5' deletions were made.

Determination of the Minimum Sequence Required for Termination. To more precisely define the termination signal, a set of 15 additional duplex oligonucleotides with *Acc* I and *Eco*RI ends were inserted into the VGF gene in place of AATTCTTAT. In each case the oligonucleotide was bounded by *Acc* I (upstream) and *Eco*RI (downstream) linkers. The formation of terminated and read-through transcripts was determined by autoradiography of polyacrylamide gels (Fig. 3A and B) and quantitated by densitometry (Table 1). We found that deletion of the first of two adenosine residues preceding the thymidine had little effect on termination (Fig. 3A, lane 3). Termination still occurred when the next adenosine was removed but the efficiency appeared to be somewhat lower (lane 4). By contrast, deletion of one thymidine out of the block of five thymidines (lane 7) or replacement of the middle thymidine with an adenosine (lane 12) resulted in

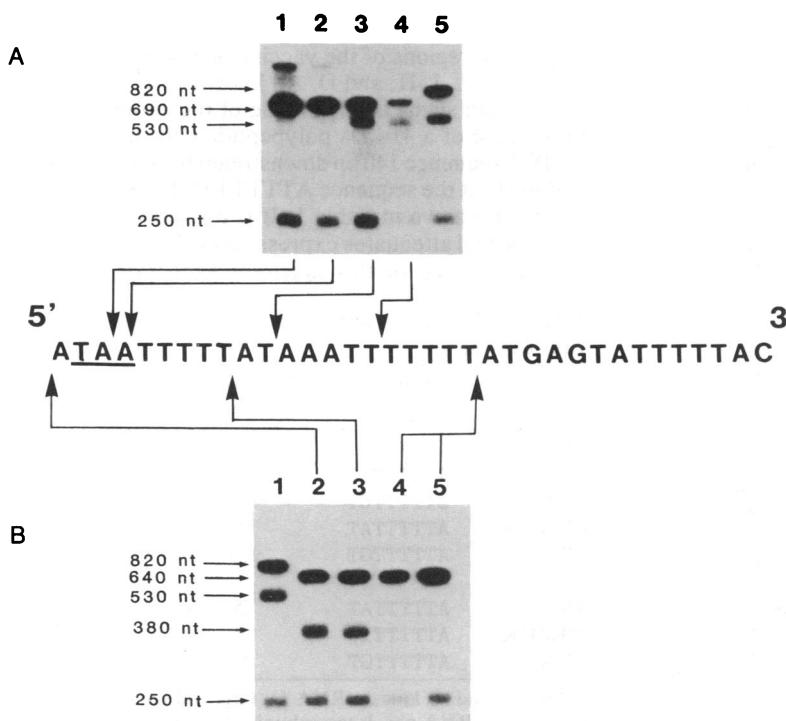


FIG. 2. Effects of deletions on transcription termination. The nucleotide sequence shown starts 1 nt before the VGF stop codon (underlined) and continues downstream. (A) *In vitro* transcription was carried out using purified *Nde* I fragments containing the entire VGF gene (lane 5) or the VGF gene containing progressive 3'→5' deletions (in order, lanes 4 to 1). The extent of deletion for each template is indicated. The transcription products were analyzed by polyacrylamide gel electrophoresis and autoradiography. For the undeleted template, the VGF terminated and run-off transcripts were 530 and 820 nt, and the run-off gene X transcript was 250 nt. For the deleted templates, the transcript sizes were the same except that the VGF run-off transcripts were ≈690 nt. (B) *In vitro* transcription was carried out using purified *Nde* I fragments containing the entire VGF gene (lane 1) or the VGF gene containing progressive 5'→3' deletions (lanes 2 to 5). For the deleted templates, the run-off transcripts were ≈640 nt, and the terminated transcripts were 380 nt.

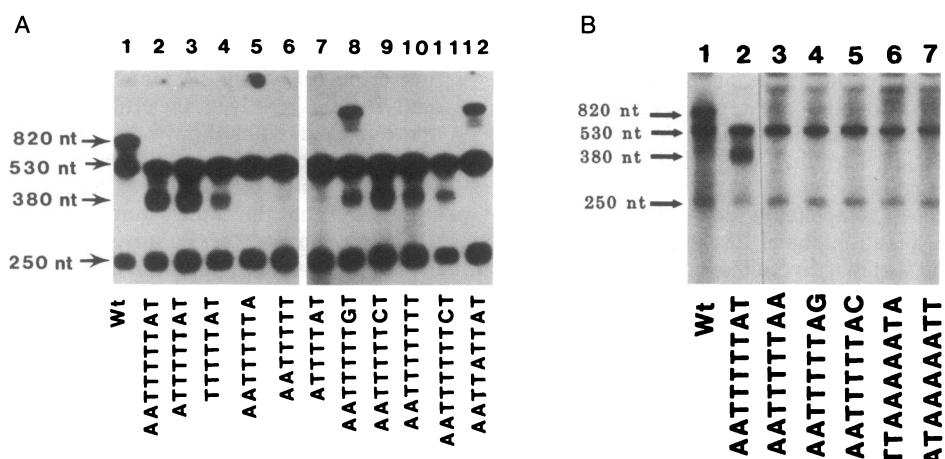


FIG. 3. Replacement of the termination signal with synthetic oligonucleotides. Double-stranded oligonucleotides with *Acc* I and *Eco* RI linkers were used to replace the 300-bp 3'-terminal segment of the VGF gene in plasmid pLY1. The *Nde* I fragment containing the promoter and putative terminator region of each plasmid was used as a template for transcription. The RNA products were analyzed by polyacrylamide gel electrophoresis and autoradiography. The terminated and run-off transcripts of the wild-type (Wt) template were 530 and 820 nt, respectively. The terminated and run-off transcripts of the mutant templates were 380 and 530 nt, respectively. The 250 nt RNA is the run-off product of the truncated X gene. The sequences of the synthetic oligonucleotides (without the *Acc* I and *Eco* RI linkers) are shown below each lane.

the formation of read-through transcripts exclusively. Deletion of the rightmost thymidine (lane 5) or replacement of that thymidine with adenosine, guanosine, or cytidine (Fig. 3B, lanes 3, 4, and 5, respectively) also abolished termination. Substitution of other nucleotides for the adenosine of TT-TTTAT, however, only lowered the efficiency of termination (Fig. 3A, lanes 8–10). Neither the complementary sequence TTAAAAATA (Fig. 3B, lane 6) nor the backwards complementary sequence ATAAAAATT (Fig. 3B, lane 7) supported termination. We concluded that the minimum termination signal is TTTTTNT, but that ATTTTTAT is more efficient.

With the information obtained by oligonucleotide mutagenesis, we inspected the sequence downstream of the VGF stop codon for termination signals. The first block of thymidines following the VGF gene contains the termination signal ATTTTTAT, and the second contains two potentially overlapping signals ATTTTTTT and TTTTTAT. The closest approximation to a termination signal in the third block of

thymidines, ATTTTTAC, is missing the required terminal thymidine and should, therefore, be ineffective. This interpretation precisely matches the conclusions from deletion mutagenesis, that independent termination signals are present in the first two sets of thymidines but not in the third.

Occurrence of Termination Signals in Other Vaccinia Early Genes. If the sequence TTTTTNT is extensively used as a signal for transcriptional termination, it should occur frequently near the ends but only rarely if at all in the middle of early genes. Accordingly, we first examined the sequences of those early genes for which 3'-end transcriptional mapping has been reported. We found that in each case a consensus termination signal was 20–50 bp before the 3' end of the mRNA (Table 2). In all but one case, an adenosine preceded the run of thymidines consistent with the higher efficiency of termination that was found with this arrangement *in vitro*. The N position was variously filled with adenosine, thymidine, cytidine, or guanosine. To extend this analysis, a computer search for TTTTTNT was made in the long sequenced regions of the vaccinia genome including *Hind* III fragments L, J, H, and D. All known early genes lacked the termination signal in the middle of the coding region except in the case of a 41-kDa polypeptide (19) that has an ATTT-TTTCT sequence 140 bp downstream from the ATG initiation codon. That the sequence ATTTTTCT can serve as a signal *in vitro* is shown in Table 1. It is possible that the presence of this signal attenuates expression of the gene. To determine the significance of this apparent exception, however, the

Table 1. Sequence requirements for termination

Oligonucleotide			Relative termination efficiency, %
<i>Acc</i> I	Signal	<i>Eco</i> RI	
GTAGAC-AATTTTTAT-GAATTC			94
GTAGAC-ATTTTTAT-GAATTC			82
GTAGAC-TTTTTAT-GAATTC			42
GTAGAC-AATT-TTAT-GAATTC			ND
GTAGAC-AATTATTAT-GAATTC			ND
GTAGAC-AATTTTTGT-GAATTC			46
GTAGAC-AATTTTTCT-GAATTC			72
GTAGAC-AATTTTTT-GAATTC			57
GTAGAC-AATTTTTA-GAATTC			ND
GTAGAC-AATTTTT—GAATTC			ND
GTAGAC-AATTTTTAA-GAATTC			ND
GTAGAC-AATTTTTAC-GAATTC			ND
GTAGAC-AATTTTTAG-GAATTC			ND
GTAGAC-AATTTTTCT-GAATTC			49
GTAGAC-TTAAAAATA-GAATTC			ND
GTAGAC-ATAAAAATT-GAATTC			ND

For clarity, portions of the sequence corresponding to the *Acc* I site, the signal site, and *Eco* RI site are separated by dashes. Termination efficiency was determined from the densitometric ratio of terminated/run-off transcripts. Termination efficiencies are relative to that obtained with the natural tandem termination sequence. ND, terminated band not detected.

Table 2. Presence of termination signal in 3' region of vaccinia virus early genes

Gene	Termination sequence	Nucleotides from 3' end	Ref.
VGF	ATTTTTAT	35	15
	ATTTTTT		
7.5K	CTTTTTCT	30	16
RNA pol	ATTTTTAT	25	17
42K	ATTTTTGT	35	15 and unpublished observations
TK	ATTTTTAT	20	18
22K/41K	ATTTTTT	30	10, 19
RNA G	ATTTTTGT	50	17, 20

TK, thymidine kinase; RNA G, large subunit of RNA guanyltransferase; RNA pol, large subunit of RNA polymerase.

sequence needs to be confirmed and the *in vivo* transcripts further analyzed.

In marked contrast to the results obtained with early genes, the TTTTTNT consensus sequence was frequently found in the coding region of known late genes and genes suspected of being late because of the characteristic TAAATG sequence at the start of the open reading frame (18, 21–23).

DISCUSSION

Development of a template-dependent *in vitro* transcription system from vaccinia virions has allowed us to define the signal used for termination of transcription of early vaccinia virus genes. The apparent absence of 3' processing, which may be an adaptation to the intra-virion or intra-cytoplasmic site of transcription, simplified our task. The sequence TTTTTNT is sufficient to signal termination, which occurs about 50 nt downstream. Termination appears to be most efficient, however, when N is adenosine and an adenosine precedes the thymidine cluster. We have demonstrated (9) that the sequence at which termination occurs is not specific since vector DNA replaced vaccinia DNA without noticeable effect (9). In accord with this, any RNA with a free 3' end may be polyadenylated (9).

The significance of these *in vitro* studies can be inferred by comparison of mRNAs made *in vitro* and *in vivo* and by examination of early gene sequences. Studies have indicated (9, 13) that the 3' ends of the *in vitro* transcripts correspond to those of VGF mRNA made *in vivo* suggesting use of the same termination mechanism. In addition, the consensus termination sequence is found just upstream of the 3' end of all early mRNAs examined (Table 2) and rarely in coding segments of early genes. Since termination occurs downstream of the termination signal, the presence of the consensus sequence near the end of the coding segment would presumably not interfere with transcription of the entire gene. Recognition of the termination signal can explain some previously puzzling data. For example, the absence of a termination sequence at the end of one early gene encoding a 41-kDa polypeptide (19) indicates why the 3' end of that mRNA is coterminous with the gene immediately downstream (24) that does have a termination signal. There are other examples of mRNAs, such as the thymidine kinase, in which the majority of species terminates after the open reading frame and a minor amount further downstream indicating this mechanism is not 100% efficient even *in vivo* (24). Although tandem copies of the consensus sequence occur occasionally, single copies are more usually found. The frequent finding of an adenosine preceding the thymidine cluster suggests that this arrangement increases the efficiency of termination *in vivo* as well as *in vitro*.

Studies have indicated that vaccinia virus late mRNAs have extremely heterogeneous 3' ends that can run for thousands of nucleotides past the first coding sequence. Such a situation could occur if late genes have no termination signals or if alterations in the transcription mechanism occurred. The latter explanation seems the more important since TTTTTNT sequences were consistently found within late coding regions where they would interfere with transcription if recognized.

Assuming that the vaccinia virus transcription system is derived from or shares its origin with a eukaryotic source, we might expect to find related sequences used for mRNA termination in eukaryotes. In fact, there is preliminary evidence that ATTTTTAT is part of a longer RNA polymerase II termination signal in yeast (4) and in mammals (6) and may regulate termination in an *in vitro* system derived from HeLa cells (25). Thymidine-rich clusters also form part of a putative tripartite sequence proposed for sea urchin histone mRNA termination (5).

Termination by vaccinia virus RNA polymerase appears to be quite different from that occurring with RNA polymerase III, although the latter also uses thymidine-rich sequences. The purified *Xenopus laevis* enzyme terminates transcription of 5S genes within a cluster of four or more consecutive thymidine residues that is surrounded by GC-rich sequences (26, 27). Only the sequence in the immediate vicinity of the thymidine cluster is required for termination that occurs after just two of the thymidine residues are transcribed. In contrast, the thymidine cluster forming the vaccinia virus termination signal is not surrounded by GC-rich sequences and occurs well upstream of the termination site.

Escherichia coli ρ factor-mediated termination (28) is a possible model for vaccinia virus. It has been suggested that ρ factor binds to AU-rich regions of nascent transcripts and is propelled along the RNA by its nucleoside triphosphatase activity until it contacts a paused RNA polymerase. At that point, the RNA is released resulting in termination. Vaccinia virus cores contain two nucleic acid-dependent nucleoside triphosphatases of unknown function (29).

There are a number of important but still unanswered questions regarding vaccinia virus transcription termination. For example, is the actual signal in DNA or RNA? Is the signal recognized by the RNA polymerase itself or by a special termination factor?

We thank Steve Broyles for providing RNA polymerase preparations and advice and Joseph Baldick for synthesizing oligonucleotides.

- Citron, B., Falck-Pedersen, E., Salditt-Georgieff, M. & Darnell, J. E. (1984) *Nucleic Acids Res.* **12**, 8723–8731.
- Falck-Pedersen, E., Logan, J., Shenk, T. & Darnell, J. E., Jr. (1985) *Cell* **40**, 897–905.
- Hagenbuchle, O., Wellauer, P. K., Cribbs, D. C. & Schibbler, U. (1984) *Cell* **38**, 737–744.
- Henikoff, S. & Cohen, E. H. (1984) *Mol. Cell. Biol.* **4**, 1515–1520.
- Johnson, M. R., Norman, C., Reeve, M. A., Scully, J. & Proudfoot, N. A. (1986) *Mol. Cell. Biol.* **6**, 4008–4018.
- Sato, K., Ito, R., Baek, K. H. & Agarwal, K. (1986) *Mol. Cell. Biol.* **6**, 1032–1043.
- Yarger, J. G., Armilei, G. & Gorman, M. C. (1986) *Mol. Cell. Biol.* **6**, 1095–1101.
- Zaret, K. S. & Sherman, F. (1982) *Cell* **28**, 563–573.
- Rohrmann, G., Yuen, L. & Moss, B. (1986) *Cell* **46**, 1029–1035.
- Broyles, S. S. & Moss, B. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3141–3145.
- Moss, B. (1985) in *Virology*, eds. Fields, B. N., Knipe, D. M., Chanock, R. M., Melnick, J., Roizman, B. & Shope, R. (Raven, New York), pp. 658–704.
- Rohrmann, G. & Moss, B. (1985) *J. Virol.* **56**, 349–355.
- Yuen, L. & Moss, B. (1986) *J. Virol.* **60**, 320–323.
- Broyles, S. S. & Moss, B. (1987) *Mol. Cell. Biol.* **7**, 7–14.
- Venkatesan, S., Gershowitz, A. & Moss, B. (1982) *J. Virol.* **44**, 637–646.
- Venkatesan, S., Baroudy, B. M. & Moss, B. (1981) *Cell* **25**, 805–813.
- Morgan, J. R., Cohen, L. K. & Roberts, B. E. (1984) *J. Virol.* **52**, 206–214.
- Weir, J. P. & Moss, B. (1983) *J. Virol.* **46**, 530–537.
- Plucienniczak, A., Schroeder, E., Zettlemoissl, G. & Streeck, R. E. (1985) *Nucleic Acids Res.* **13**, 985–998.
- Niles, E. G., Condit, R. C., Caro, P., Davidson, K., Matusick, L. & Seto, J. (1986) *Virology* **153**, 96–112.
- Rosel, J. & Moss, B. (1985) *J. Virol.* **56**, 830–838.
- Rosel, J. L., Earl, P. L., Weir, J. P. & Moss, B. (1986) *J. Virol.* **60**, 436–449.
- Wittek, R., Hanggi, M. & Hiller, G. (1984) *J. Virol.* **49**, 371–378.
- Bajszar, G., Wittek, R., Weir, J. P. & Moss, B. (1983) *J. Virol.* **45**, 62–72.
- Baek, K. H., Sato, K., Ito, R. & Agarwal, K. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7623–7627.
- Bogenhagen, D. F. & Brown, D. D. (1981) *Cell* **24**, 261–270.
- Cozzarelli, N. R., Gerrard, S. P., Schlissel, M., Brown, D. D. & Bogenhagen, D. F. (1983) *Cell* **34**, 829–835.
- Holmes, W. M., Platt, T. & Rosenberg, M. (1983) *Cell* **32**, 1029–1032.
- Paoletti, E. & Moss, B. (1974) *J. Biol. Chem.* **249**, 3281–3286.