The EMBO Journal vol.7 no.4 pp.1183 - 1190, 1988

Vaccinia virus late transcripts generated *in vitro* have a poly(A) head

Beate Schwer and Hendrik G.Stunnenberg

European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, FRG

Communicated by L.Philipson

A cell free system mediating accurate transcription of vaccinia virus genes was established using lysates of cells in the late phase of infection. Vaccinia late genes are faithfully transcribed in this extract whereas cellular *pol* II and *pol* III promoters are not recognized. The late viral transcripts contain a poly(A) head of ~35 nt at the 5′ end which is not co-linearly encoded in the externally added template. The transcripts obtained *in vitro* are indistinguishable from the mature *in vivo* RNAs. The poly(A) head is synthesized *de novo* and its formation appears to be directly coupled to the transcription of the gene. The synthesis of the poly(A) head transcripts *in vitro* is consistent with a proposed slippage model.

Key words: vaccinia virus/*in vitro* transcription/poly(A) head/cap analogue/slippage

Introduction

The rate and the temporal order of gene expression in prokaryotes and eukaryotes can be regulated at the transcriptional and/or translational level. Gene expression is very often regulated at the transcriptional level through the cooperative action of several cell- and tissue-specific factors (activators and repressors). Some viruses and parasites have developed 'exotic' mechanisms, e.g. primed initiation in corona virus (Spaan *et al.*, 1982; Makino *et al.*, 1986) and influenza virus (Beaton and Krug, 1981), *trans*-splicing of glycoprotein RNAs in trypanosomes (Murphy *et al.*, 1986; Sutton and Boothroyd, 1986) or interspersed addition of U-residues in mitochondrial transcripts in trypanosomes (Benne *et al.*, 1986; Feagin *et al.*, 1987).

Vaccinia virus regulates its gene expression in a temporal fashion in the cell cytoplasm. The activation of early genes prior to viral DNA replication is accompanied by a shut-off of host transcription and translation and the genes expressed after DNA replication are referred to as late (Moss, 1985). Vaccinia virus uses an exceptional mechanism to regulate its gene expression. Activation of the late genes is not merely the consequence of the presence or absence of activator or repressor molecules but rather the mode of transcription changes upon switching from early to late. While early genes are transcribed from a contiguous DNA segment (Venkatesan *et al.*, 1981; Yuen and Moss, 1986), late transcripts are discontinuously synthesized and include a poly(A) leader upstream of the coding part of the gene which is not colinearly encoded in the genome (Bertholet *et al.*, 1987; Schwer *et al.*, 1987). Bertholet and collaborators isolated cDNA clones from the 11K gene having 5′ poly(A) stretches of heterogeneous length and some of the cDNAs contain additional RNA sequences upstream of the poly(A) tract (Bertholet *et al.*, 1987). Our results revealed that late transcripts have an almost discrete and homopolymeric stretch of ~35 A-residues which is preceded by a m^7G cap structure (Schwer *et al.*, 1987). Both groups mapped the junction between the poly(A) head and the coding body of the messenger within a TAAATG sequence motif (Bertholet *et al.*, 1987; Schwer *et al.*, 1987). This sequence motif is highly conserved and essential for transcription of late genes (Rosel *et al.*, 1986; Hänggi *et al.*, 1986). The fact that late transcripts have a 5′ poly(A) stretch raises intriguing questions about its biological significance and the nature and specificity of the vaccinia virus RNA polymerase in the early and late stage of the infection cycle.

The development of a cell-free system mediating accurate transcription on exogenous DNA templates *in vitro* is a prerequisite to study transcription regulation. Soluble transcription systems specific for early vaccinia genes have been prepared from purified virions (Golini and Kates, 1985; Rohrmann and Moss, 1985) and from cells in the early phase of virus infection (Foglesong, 1985; Puckett and Moss, 1983). We established an *in vitro* transcription system from cells in the late phase of infection, which preferentially recognizes vaccinia virus late regulatory sequences. Transcripts made in this *in vitro* extract were analysed for the presence of a 5′ terminal poly(A) stretch and the putative mechanism involved in the generation of the late transcripts is discussed.

Results

In vitro transcripts have a poly(A) head

The development of a cell-free system specific for vaccinia transcription might be crucial to clarify the mechanism involved in the synthesis of the poly(A) head mRNA. We have, therefore, prepared whole cell extracts from HeLa suspension cultures infected with vaccinia virus harvested in the late phase of infection. Virus specific transcription was measured in a run-off assay using a chimaeric gene construct (11K−DHFR) consisting of the regulatory sequences of the 11K late promoter (from position −100 to +8), necessary for faithful expression in the late phase, and the coding sequences of the mouse dihydrofolate reductase (DHFR) gene (Hänggi *et al.*, 1986; Schwer *et al.*, 1987).

Run-off transcripts of ~385 and 400 nt in length are synthesized *in vitro* from the chimaeric gene depending on the position of the truncation (Figure 1A, lanes 2 and 4). The size of the run-off transcripts is ~35 nt longer than the distance between the position of the truncation and the TAAATG-motif at the 5′ border of sequences colinearly encoded in the DNA; this was revealed by S1 mapping. Transcripts of heterogeneous length detectable in the upper part of the gel are nonspecific and not promoter dependent.

They are also synthesized in the absence of exogenous DNA although at lower levels (Figure 1A, lane 1) or if a DHFR construct lacking the 11K promoter is used (Figure 1A, lane 3).

The 5′ termini of the *in vitro* transcripts were analysed by S1 mapping. S1 protection experiments using a genomic DNA fragment result in the generation of fragments of 460 and 350 nt in length (Figure 1C, lane 1). The latter fragment represents protection of DHFR sequences up to position +1 (the A of AUG is referred to as position +1). The fragment of 460 nt (indicated by an arrow head) results from protection by readthrough transcripts initiated on plasmid sequences. The presence of a non-contiguous poly(A) head could be established by S1 mapping with a probe derived from an artificial DNA fragment in which the promoter sequences upstream of position +1 were replaced by a synthetic poly(A) tract of 80 A-residues (schematically drawn in Figure 1D). The S1 protected fragment of 385 nt can only arise from protection of DHFR transcripts which are preceded by a poly(A) head of ~35 A-residues. The faster migrating fragment (indicated by the arrow head) corresponds to sequences which are only homologous with the DHFR sequences. They might represent readthrough transcripts which are initiated within plasmid sequences or be generated by nibbling of nuclease S1. The microheterogeneity in the length of the poly(A) tract *in vitro* is also observed *in vivo*. We conclude that the *in vitro* transcripts made in the extracts are genuine late transcripts which are indistinguishable with respect to the poly(A) head from *in vivo* mRNAs.

Optimal assay conditions

Sawadogo and Roeder (1985) have shown that a high background of nonspecific transcription can be eliminated if a G-less reporter gene is used as template and GTP is omitted from the reaction mixture. We cloned this G-less cassette in the opposite orientation downstream of position +5 of the 11K promoter, obtaining an 11K C-less reporter gene construct (11KΔC). The levels of run-off transcripts from the 11KΔC construct in the presence and absence of CTP (Figure 1B, lanes 1 and 2) or from the 11K−DHFR construct are effectively the same (Figure 1B, lane 3). It can be concluded that transcripts from the 11KΔC chimaeric gene are accurately initiated at promoter sequences upstream of the C-less cassette in the absence of CTP whereas the non-specific background transcription is eliminated.

Transcription assays are routinely performed using 4−6 μl of extract in a final volume of 10 μl depending on the preparation (data not shown). The optimal concentrations of Mg^{2+} and KCl in our assay are 2.5 and 50 mM, respectively (Figure 2A). Mg^{2+} can be substituted by Mn^{2+} in the reaction without significant loss of polymerase activity; specific transcription in the presence of Mn^{2+} is only obtained in a narrow concentration range (5−10 mM) with the optimum at 7.5 mM (data not shown). Similar assay conditions have been established for the early specific vaccinia RNA polymerase in extracts derived from purified virions (Golini and Kates, 1985) or from infected cells harvested at the early stage of infection (Puckett and Moss, 1983).

Maximal transcription is obtained at a DNA concentration of 100 μg/ml using truncated (Figure 2B) or supercoiled templates (data not shown). Linear templates, however,

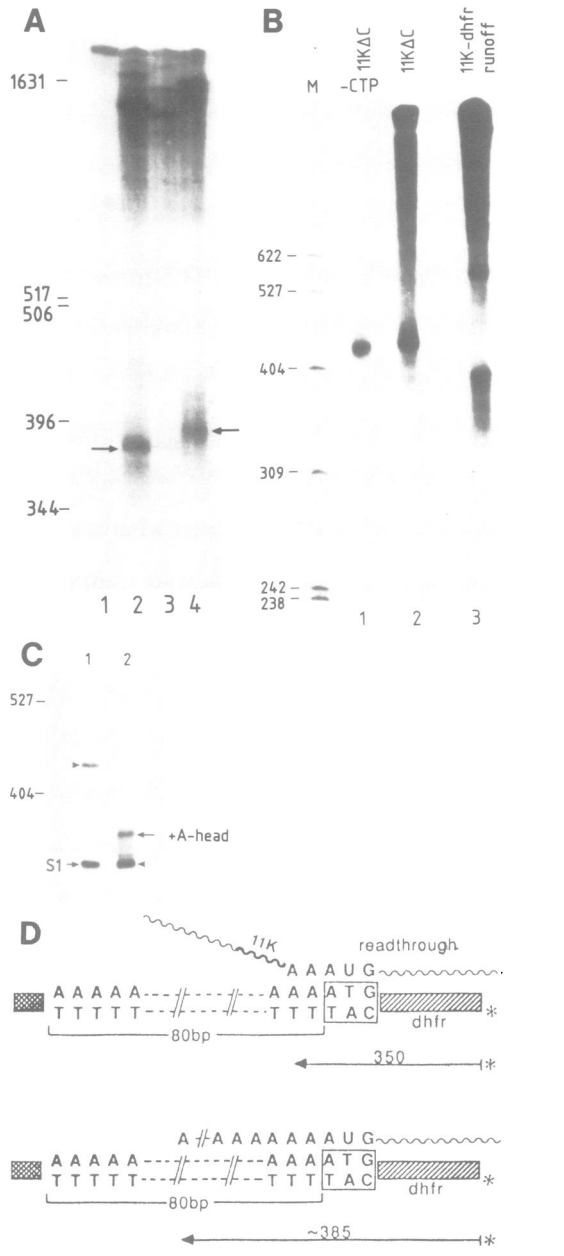

Fig. 1. *In vitro* transcription of the vaccinia 11K promoter using an extract from virus-infected HeLa cells. (**A**) Transcripts synthesized in the absence of exogenous DNA template (**lane 1**) or in the presence of 100 μg/ml of the 11K−DHFR template truncated at position +350 (**lane 2**) or +365 (**lane 4**). RNA from a DHFR construct without the 11K promoter is shown in **lane 3**. The A-residue of the ATG within the conserved TAAATG motif of the promoter is referred to as position +1. The arrows indicate the position of the main transcripts. This extract has not been chromatographed over DEAE−Sephacel. (M): ^{32}P-labelled *Hin*fI digested pBR322 size marker. (**B**) Run-off transcripts synthesized from the 11KΔC template truncated at position +390 (**lanes 1** and **2**) or 11K−DHFR template truncated at +350 (**lane 3**) using standard assay conditions. Transcription was performed in the presence of all four ribonucleotides (**lanes 2** and **3**) and in the absence of CTP (lane 1). (M) ^{32}P-labelled *Hpa*II digested pBR322 size marker. (**C**) S1 mapping of the *in vitro* transcripts from the 11K−DHFR template using 20 U of nuclease S1 enzyme per assay and a 5′ end-labelled genomic DNA probe from −115 to +350 (**lane 1**) or an artificial DNA probe containing 80 A-residues and additional plasmid sequences upstream of DHFR sequences from position +1 to +350 (**lane 2**). (i) Input S1 probe; (>) readthrough transcripts. The arrow indicate the run-off transcripts with a 5′ terminal poly(A) head. (M) ^{32}P-labelled *Hin*fI digested pBR322 size marker. (**D**) Schematic drawing of S1 mapping using a DNA fragment in which promoter sequences upstream of the ATG were replaced by an A-tract.

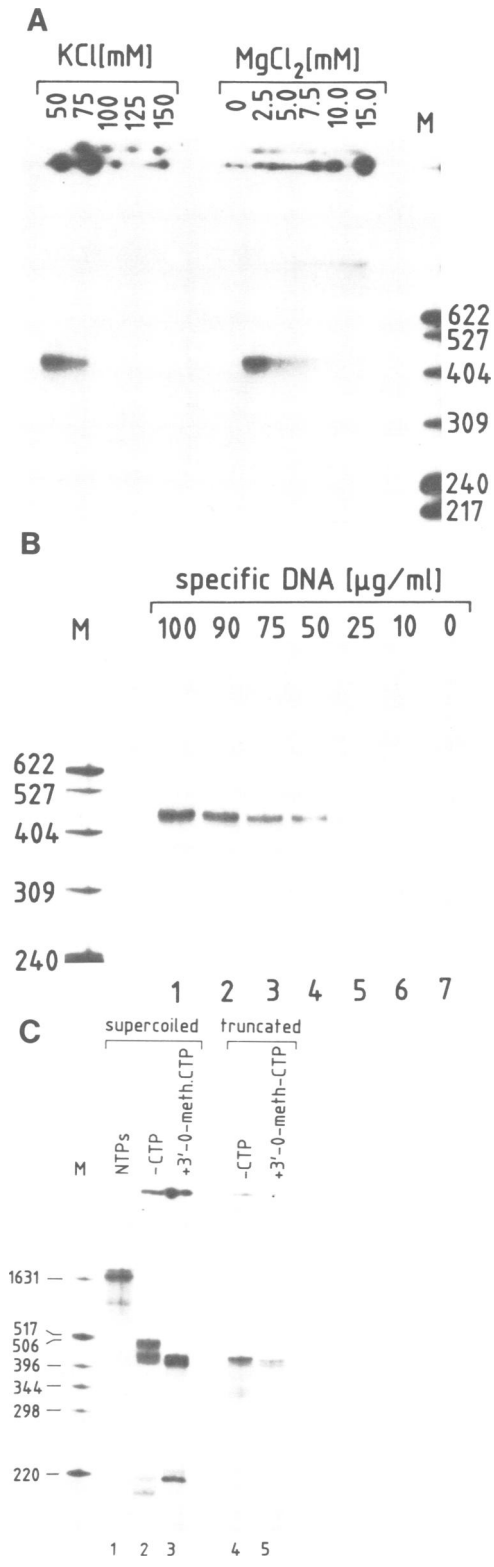

Fig. 2. Determination of the optimal assay conditions. (**A**) The effect of increasing concentrations of Mg^{2+} and KCl on the RNA synthesis in the presence of 100 µg/ml of the truncated 11KΔC template. 0 mM $MgCl_2$ was obtained by adding 2.5 mM EDTA to the reaction mixture. (M) ^{32}P-labelled *Hpa*II digested pBR322 size marker. (**B**) The effect of increasing concentrations of the 11KΔC template truncated at position +390 on the RNA synthesis. (M) ^{32}P-labelled *Hpa*II digested pBR322 size marker. (**C**) Run-off transcription using standard buffer conditions and 100 µg/ml of supercoiled (**lanes 1–3**) or truncated (**lanes 4** and **5**) 11KΔC template in the presence and absence of CTP or in the presence of the chain-terminator 3′-*o*-methyl-CTP. (M) ^{32}P-labelled *Hin*fI digested pBR322 size marker.

appear to be 3- to 5-fold less efficiently transcribed than supercoiled templates (Figure 2C). This result could be confirmed by kinetically measuring TCA precipitable material (data not shown). Transcription from the supercoiled 11KΔC template in the presence of all 4 rNTPs yields long transcripts due to the absence of specific termination (Figure 2C lane 1). Omitting CTP from the reaction mixture results in the synthesis of two major transcripts ~420 and 480 nt in length (lane 2). The longer transcripts are terminated in a C-rich region within plasmid sequences downstream of the C-less cassette, which can be explained by the presence of low levels of CTP in our extract. Addition of 3′-*o*-methyl-CTP to the reaction mixture results in termination of RNA synthesis at the 3′ border of the C-less cassette (lane 3). A relatively high proportion of the transcripts terminates within the reporter gene which has not been further investigated.

Transcription is dependent on vaccinia promoters

The host transcription and translation is reported to be shut off upon infection of the cells with vaccinia virus (Moss, 1985). We have investigated whether any residual host polymerase activity could be detected in the extract which could account for the synthesis of the poly(A) head transcripts in our *in vitro* transcription system. Whole cell extracts have been prepared from mock-infected and virus-infected HeLa cells using the same procedures. HeLa RNA polymerase II and III activity are readily detectable in extracts from mock-infected cells (Figure 3A) using the adenovirus major late promoter in front of the G-less cassette (Sawadogo and Roeder, 1985) and a 5S maxi gene from *Xenopus* (Sakonju *et al.*, 1980). Cellular polymerase activity is not detectable anymore in extracts derived from vaccinia-infected cells (Figure 3B). We cannot detect specific transcripts derived from the 11KΔC template in extracts from mock-infected cells but run-off transcripts are synthesized in extracts from virus-infected cells. Vaccinia RNA polymerase activity is insensitive to α-amanitin even at the highest concentration tested whereas host RNA polymerase II and III activities are sensitive (Figure 3B, lanes 6–8, and Figure 3A, lanes 1–5, respectively). Early vaccinia promoters are not transcribed in extracts from mock-infected cells (data not shown). Low levels of early transcripts are detectable in extracts from cells in the late phase of infection, which is probably due to co-extraction of early specific polymerase enzyme from mature virions present within the cells at this stage of infection. The results demonstrate that the extract from infected cells is highly specific for vaccinia promoters with a marked preference for the late class of promoters and that host RNA polymerase activity is absent.

Time course of RNA synthesis

We have investigated whether a precursor RNA transcript is synthesized in our extract which would be indicative of a post-transcriptional mechanism of A-addition. We have therefore analysed run-off transcripts from the 11KΔC construct generated at different incubation times in the absence of CTP. Transcripts with a 5′ terminal poly(A) head are detectable after a short lag period of a few minutes (2–3 min on long exposures) and the rate of synthesis levels off after 30–45 min (Figure 4). We do not detect shorter or longer run-off transcripts without a poly(A) tract even at the earliest time points and on very long exposures. This indicates that the poly(A) addition does not occur post-transcriptionally; only if the poly(A) synthesis is very fast

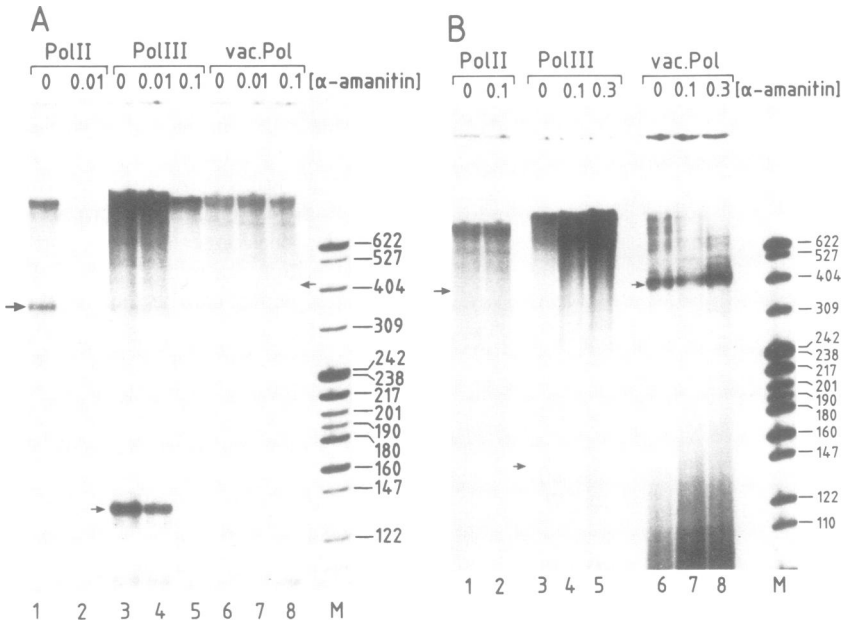

Fig. 3. RNA polymerase activities in extracts from mock-infected (**A**) or vaccinia virus-infected HeLa cell suspensions (**B**). The extracts were prepared as described in Materials and methods and the run-off transcription was performed in standard assay conditions and 100 μg/ml of DNA template. The HeLa RNA polymerase II and III activities were determined using pML(C2AT)19, truncated at position +390, in the absence of GTP (Sawadogo and Roeder, 1985) and the 5S maxigene of *Xenopus* (Sakonju *et al.*, 1980), respectively. Vaccinia late specific RNA polymerase activity was determined using the 11KΔC template, truncated at position. +390, in the absence of CTP. Different concentrations of the inhibitor α-amanitin were added to the reaction mixture as indicated (μg/ml). (M) ^{32}P-labelled *Hpa*II digested pBR322 size marker. The arrows indicate the positions of the respective run-off transcripts.

compared to the rate of transcription or if it is added to a nascent RNA chain we would not be able to detect a precursor in this type of experiment.

Effect of ATP

Next we investigated whether the synthesis of the poly(A) head is coupled with the transcription of the reporter gene or whether a *de novo* synthesized poly(A) RNA or an RNA molecule present in the extract (e.g. in the form of an RNP particle) acts as a primer of transcription. We have therefore constructed an A-less reporter gene (ΔA) downstream of the 11K promoter which could potentially allow an uncoupling of the synthesis of the poly(A) head from the synthesis of the reporter gene. The transcription of the 11KΔA construct truncated at position +275 in the presence of increasing amounts of ATP is shown in Figure 5. Creatine phosphate (10 mM) is added to the transcription reactions to regenerate the ATP. Run-off transcripts with a poly(A) head are synthesized in the presence of 0.5−5 mM ATP while higher concentrations inhibit the transcription activity which might be due to complexing Mg^{2+} with ATP. An unanticipated effect is obtained at low ATP concentrations: transcription is not merely reduced but the run-off transcripts are significantly shorter and more variable in length. Further decrease of the ATP concentration (<50 μM) abolishes transcription of specific RNAs but not the high background of nonspecific labelling. The same result is obtained if transcription is performed at low ATP concentrations and in the absence of CTP using the 11KΔC construct as template (Figure 5B). We have confirmed by 'artificial' S1 mapping that the variation in the length of the run-off transcripts is the result of a variable length of the A-head (Figure 5C) using the same approach as schematically indicated in Figure 1D. 5′ terminal poly(A)

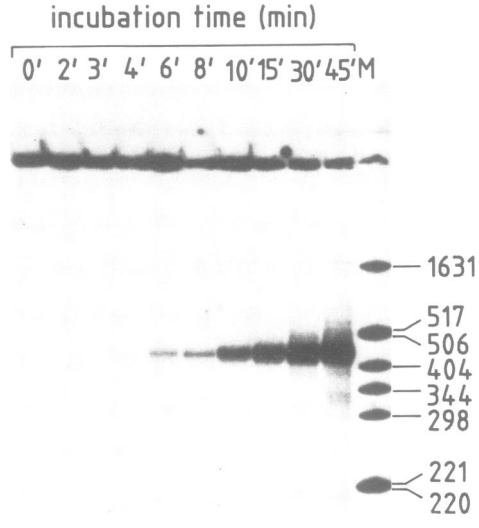

Fig. 4. Time course of RNA synthesis *in vitro*. Run-off transcription was performed in the absence of CTP using standard reaction conditions and 100 μg/ml of the 11KΔC template truncated at position +390. Aliquots were taken at different times of incubation as indicated at the top of the gel. (M) ^{32}P-labelled *Hin*fI digested pBR322 size marker.

sequences of ~35 nt are covalently linked to the RNA transcripts if 0.5−1.0 mM of exogenous ATP is added to the reaction (Figure 5C, lanes 5 and 7). Protected fragments migrating at the position of the S1 'start site' represent background transcripts initiated at plasmid sequences. Increasing the concentration of nuclease S1 diminishes the protected fragment obtained with poly(A) head transcripts probably due to nibbling of the enzyme at the unstable

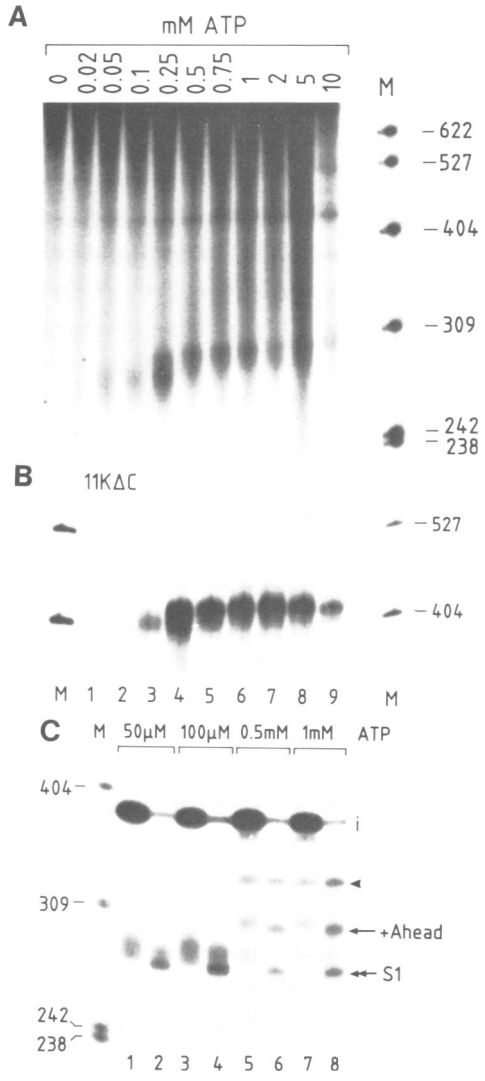

Fig. 5. The effect of the ATP concentration on the synthesis of run-off transcripts. (**A**) transcripts synthesized from the 11KΔA template truncated at position +275. Transcription was performed at different ATP concentrations, as indicated in mM ATP. (M) ^{32}P-labelled *Hpa*II digested pBR322 size marker. (**B**) ATP dependence of transcription of the 11KΔC template truncated at position +390. **Lanes 1–9**: 0, 0.05, 0.1, 0.25, 0.5, 1, 2, 5 and 10 mM ATP. (**C**) Nuclease S1 protection of *in vitro* transcripts synthesized from supercoiled 11KΔA template at different ATP concentrations. A 5′ end-labelled probe containing the A-free cassette and a stretch of 65 A residues plus additional plasmid sequences upstream of the ATG was used. **Lanes 1, 3, 5** and **7**: protected DNA fragments generated with 5 U of nuclease S1. **Lanes 2, 4, 6** and **8**: S1 mapping with 50 U of the enzyme. (M) ^{32}P-labelled *Hpa*II digested pBR322 size marker. (>) full-length protected 65 A residues.

poly(A)/poly(dT) termini (Figure 5C, lanes 6 and 8). The poly(A) stretches obtained at low levels of exogenous ATP are significantly shorter and very heterogeneous in length (Figure 5C, lanes 1 and 3). Again, a high level of protected fragments migrating at the S1 position is generated at higher nuclease S1 concentrations (Figure 5C, lanes 2 and 4). The full-length protection of the T-stretch in the S1 probe (Figure 5C, lanes 5–8 indicated by an arrow head) is only obtained with the 11KΔA template and seems to depend on gene internal sequences. Neither the 11K-DHFR construct (Figure 1C) nor the 11K authentic gene (data not shown) give rise

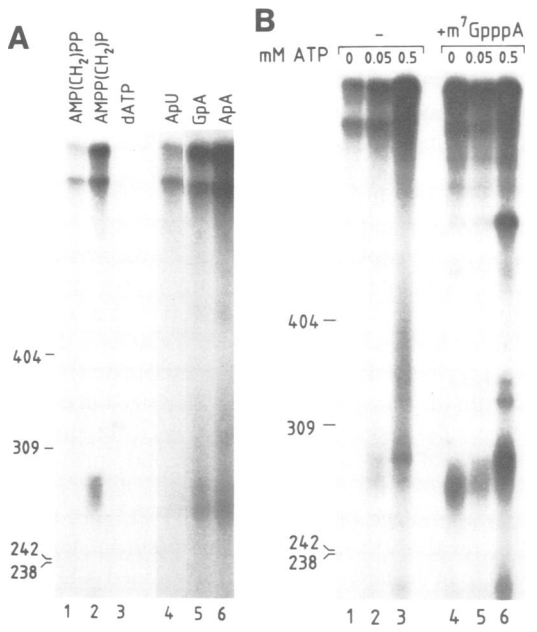

Fig. 6. Use of ATP-analogues and dinucleotides. (**A**) Run-off transcripts from the 11KΔA template synthesized in the absence of ATP. The transcription reactions were supplemented by 0.5 mM of AMP(CH$_2$)PP (**lane 1**), 0.5 mM AMPP(CH$_2$)P (**lane 2**), 0.5 mM dATP (**lane 3**) and by 0.5 mM of each of the dinucleotides: ApU (**lane 4**), GpA (**lane 5**) and ApA (**lane 6**). (M) ^{32}P-labelled *Hpa*II digested pBR322 size marker. (**B**) Transcripts synthesized from the 11KΔA construct truncated at position +265 in the absence and presence of 0.5 mM m^{7}GpppA (**lanes 1–3** resp. **4–6**). The transcription reactions were supplemented with different amounts of ATP as indicated. (M) ^{32}P-labelled *Hpa*II digested pBR322 size marker.

to high levels of transcripts with 65 or longer A-stretches. The same results have been obtained with *in vivo* RNA (Schwer *et al.*, 1987).

The gradual reduction of the length of the A-head at the 5′ end of the run-off transcripts at low ATP levels does not result in the synthesis of RNA transcripts migrating at the position of the S1 'start site' (transcripts without an A-head initiated within the TAAAT-motif). Transcripts at the S1 position would be anticipated if the A-addition is not directly coupled with the transcription of the reporter gene, e.g. in case of a post-transcriptional processing. The results, however, indicate that the poly(A) addition is directly coupled with the initiation of transcription or is occurring on nascent RNA transcripts.

Use of ATP analogues and dinucleotides

ATP might be an essential cofactor needed in the transcription reaction apart from being a substrate for the synthesis of the poly(A) head. Several structural analogues were tested for their ability to substitute ATP in the transcription reaction using 11KΔA as template. The methylene analogue AMP(CH$_2$)PP can serve as energy cofactor through hydrolysis of the $\beta-\gamma$ phosphate bond but it cannot be used in a polymerization reaction. Addition of AMP(CH$_2$)PP (0.5 mM) with a non-hydrolysable $\alpha-\beta$ phosphate bond does not result in the synthesis of specific transcripts (Figure 6A, lane 1). Furthermore, ATP cannot be substituted by dATP in the transcription reaction (Figure 6A, lane 3). The methylene analogue AMPP(CH$_2$)P with a non-hydrolysable

$\beta-\gamma$ bond give rise to synthesis of specific run-off transcripts from the 11KΔA template. The results indicate that the synthesis of the poly(A) head transcripts is not dependent on ATP as an energy source. The length of the poly(A) head synthesized in the presence of the AMPP(CH_2)P analogue is significantly shorter than that obtained in the presence of ATP. A similar effect was observed on 3′ poly(A) addition *in vitro* (Moore and Sharp, 1985).

An alternative explanation for the ATP requirement is that ATP concentrations below 50 μM might prevent the initiation of transcription. Several dinucleotide monophosphates at a final concentration of 0.5 mM were tested for their ability to serve as primers for initiation of transcription (Figure 6A). Specific run-off transcripts are not synthesized in the absence of dinucleotides and externally added ATP (Figure 6B, lane 1). Addition of ApA and GpA dinucleotide primers to the reaction results in the synthesis of specific RNA transcripts even in the absence of exogenous ATP (Figure 6A, lanes 6 and 5, respectively). The fact that poly(A) stretches are synthesized in the absence of externally added ATP indicates that a low level of ATP is present in our extract. The endogenous ATP concentration in the extract is not sufficient, however, for the synthesis of full-length poly(A) tracts of 35 nt (Figure 6A) which are only synthesized if additional ATP is added to the assay (data not shown). The dinucleotide ApU is not used by the polymerase as a primer for the transcription initiation (Figure 6A, lane 4). Addition of the cap structure analogue m^7GpppA to the reaction mixture results in an at least 10-fold increase in the synthesis of specific run-off transcripts (Figure 6B, lanes 4–6). Again, low ATP concentrations (<0.5 mM) do not give rise to transcripts with full-length poly(A) heads of 35 nt.

Discussion

The temporal regulation of gene expression during the vaccinia virus infection cycle is very complex. The switching from early to late transcription is not merely the consequence of activation or inactivation of a different class of genes but the mode of transcription in the early and late phase is different (Bertholet *et al.*, 1987; Schwer *et al.*, 1987). The early class mRNAs are transcribed from the genome in a continuously colinear fashion (Yuen and Moss, 1986). Late viral mRNAs have a leader sequence consisting of A-residues which is not co-linearly encoded in the genome and precedes the protein coding part of the RNAs (Bertholet *et al.*, 1987; Schwer *et al.*, 1987). We have demonstrated that the poly(A) leader sequence *in vivo* is of almost discrete length (~35 nt) with a m^7G cap structure at the 5′ terminus (Schwer *et al.*, 1987).

Here we have described an *in vitro* transcription system derived from vaccinia virus infected cells harvested in the late phase of viral infection. Accurate transcription which is insensitive to α-amanitin is obtained if a vaccinia promoter is added to the transcription system. Late promoters are preferentially transcribed whereas early viral genes are transcribed at very low levels (data not shown). Cellular RNA polymerase II and III activities are not detectable in this extract (Figures 1 and 3). The high level of nonspecific transcription is typical for a eukaryotic transcription extract and is due to initiation of nicks and free DNA ends (Slattery *et al.*, 1982; Manley *et al.*, 1980). This background transcription can be eliminated using a C-less reporter gene and omitting CTP from the reaction mixture (Figures 1B and 2). The run-off transcripts from the 11K late promoter synthesized in this *in vitro* extract appear to have a poly(A) sequence at their 5′ termini (Figure 1C). The length of the poly(A) tract (~35 A-residues) is identical to the poly(A) head which is found in mature, capped RNAs *in vivo* (Schwer *et al.*, 1987). Furthermore, we obtain the same microheterogeneity in the length of the poly(A) stretch *in vitro* and *in vivo*. It can therefore be concluded that all the factors and enzymes necessary for the synthesis of the poly(A) head transcripts are present in this cell free system and that the *in vitro* transcription process probably reflects that occurring *in vivo*.

Poly(A) head is synthesized de novo and apparently coupled with transcription initiation

Analysis of the transcripts synthesized at different times of incubation did not reveal the presence of precursor RNA molecules (Figure 4). Transcription assays using 11KΔA were performed in a further attempt to separate the poly(A) synthesis or addition (ATP substrate dependent) from the transcription of the reporter gene (independent of ATP as a substrate). A simple labelling of the transcripts derived from the 11KΔA template with [α-^{32}P]ATP proved to be very inefficient since high levels of hydrolysing enzymes are present in the extract and because of a high background due to '*trans*' labelling. The synthesis of transcripts with short and heterogeneous A-heads was obtained in transcription assays using the 11KΔA template in the presence of low concentrations of externally added ATP (Figure 5A). Furthermore, the length of the poly(A) head attached to the 5′ end of transcripts from the C-less cassette was also dependent on the ATP concentration (Figure 5B). The synthesis of both the poly(A) head and the main body of the RNA transcribed from the 11KΔC template in these experiments are dependent on ATP as a substrate. The run-off transcripts from the 11KΔC template synthesized in the presence of low ATP concentrations (50–250 μM) are full-length with respect RNA sequences which are encoded in the ΔC cassette. However, under these conditions, the A-heads at the 5′ termini of these transcripts are significantly shorter and more variable in length (Figure 5B). It can, therefore, be concluded that the poly(A) head is synthesized *de novo* and that the K_m of ATP for incorporation in the A-head is significantly higher than the K_m of ATP in the polymerization reaction (Figure 5). Very long poly(A) sequences were found at the 5′ end of transcripts derived from the 11KΔA template at higher ATP concentrations. We could only detect very low levels of transcripts with long A-heads *in vivo* (Schwer *et al.*, 1987) and *in vitro* with RNA derived from the authentic 11K gene (data not shown), 11K–DHFR chimaeric gene (Figure 1C) or the 11KΔC template. The synthesis of very long A-heads at the 5′ end of 11KΔA transcripts is apparently an intrinsic property of the sequence downstream of the 11K promoter.

Run-off transcripts from either the 11KΔA or the 11KΔC are not synthesized in the absence of externally added ATP (Figure 5). This is not due to the lack of ATP as an energy cofactor since ATP cannot be substituted by its structural analogue AMP(CH_2)PP (Figure 6A, lane 1). Futhermore, poly(A) head transcripts are generated if ATP is substituted by AMPP(CH_2)P. This analogue cannot function as an energy cofactor but it is used in the polymerization reaction

(Figure 6A, lane 2). Run-off transcripts containing a poly(A) tract can be generated even in the absence of exogenous ATP if dinucleotide primers such as ApA and GpA are added to the reaction mixture (Figure 6A). We have further shown that the m^7GpppA cap structure analogue gives rise to an at least 10-fold increase in the level of transcripts as compared to GpA, ApA or ATP (Figure 6). The higher transcription level obtained in the presence of the cap analogue m^7GpppA can only be attributed in part to the methylated state of the primer. The 3′ A-residue of the primer molecule appears to be essential for transcription initiation since GpppA is used by the viral polymerase for transcription initiation, although at lower efficiency, but m^7GpppG is not used at all (data not shown). It can be concluded that m^7GpppA is more efficiently used for transcription initiation as also reported for m^7GpppG with *Escherichia coli* and phage Sp6 RNA polymerases *in vitro* (Contreras *et al.*, 1982; Konarska *et al.*, 1984). The possibility that capped RNAs are more resistant to degradation cannot, however, be ruled out at this stage (Furuichi *et al.*, 1977; Shimotohno *et al.*, 1977). The fact that specific transcripts are synthesized in the presence of an adequate primer without addition of exogenous ATP indicates that low ATP concentrations (< 10 μM after extensive dialysis and DEAE−cellulose column chromatography) are present in our extract. However, the endogenous concentration of ATP is neither sufficient for the accurate initiation of transcription nor for the generation of 'full-length' poly(A) heads. The dinucleotide primer dependent initiation is in agreement with the proposed *de novo* synthesis of the poly(A) head in the extract. Furthermore it strongly suggests that the synthesis of the poly(A) head is directly coupled with the transcription of the coding part of the reporter gene.

The mechanism?

Different mechanisms are reported for the synthesis of discontinuous mRNAs. The discontinuous transcripts of trypanosomes are synthesized by a *trans*-splicing mechanism (Murphy *et al.*, 1986; Sutton and Boothroyd, 1986) which is probably also true for some of the actin mRNAs synthesized in *Caenorhabditis elegans* (Krause and Hirsh, 1987). *Trans*-splicing occurs through a mechanism of cutting and subsequent ligation of two independently synthesized precursor RNA molecules at specific and highly conserved splice-site sequences. Two characteristic features of the *trans*-splicing mechanism are not compatible with our *in vitro* results. First, the donor and the acceptor RNA molecules are synthesized independently and contain sequences which are removed upon splicing. Secondly, the ligation step involved in the splicing reaction is dependent on ATP as a cofactor. We have not been able to detect run-off transcripts containing sequences upstream of the TAAATG-motif, e.g. at early times of transcription (Figure 4) which could be splice acceptor molecules in the reaction. Furthermore, the addition of 'acceptor' molecules synthesized with T7 polymerase to the extract did not result in the addition of a poly(A) head (data not shown). Finally, we could demonstrate that the synthesis of poly(A) head transcripts is independent of ATP as an energy cofactor (Figure 6A) which rules out an RNA ligation reaction as an essential step in the poly(A) addition process.

Coronavirus (Makino *et al.*, 1986) and influenza virus (Beaton and Krug, 1981) fuse the discontinuous leader sequence to the transcripts via a primed initiation mechanism. An RNA molecule of discrete length (72 and 10−15 nt, respectively) is used by the respective polymerases as a primer to start the transcription of the gene. The observation that the synthesis of the poly(A) stretch is coupled with the transcription of the gene downstream of the TAAATG-motif is consistent with a primed initiation model. The observed variation in the length of the poly(A) stretch at the 5′ end of the transcripts obtained at low ATP concentrations (Figure 5A and B) is, however, not consistent with a primed initiation model as described for coronavirus (Makino *et al.*, 1986). Dinucleotide and m^7GpppA primer dependent transcription in the absence of exogenous ATP is not compatible with the *in vitro* data for influenza virus (Honda *et al.*, 1986). If an influenza type mechanism applied for the *in vitro* synthesis of poly(A) heads of vaccinia late transcripts it would be anticipated that the majority of the transcripts synthesized in the absence of exogenous ATP would lack an A-tract and would therefore migrate at the S1 position. We only obtain, however, run-off transcripts with an A-tract which is, depending on the ATP concentration, more variable in length.

Proposal of a slippage mechanism

Our results are consistent with a slippage mechanism as postulated for T4 transcription *in vitro* (Kassavetis *et al.*, 1986). Kassavetis *et al.* proposed a hypothetical event which might occur if transcription initiation starts on three A-residues in a row positioned at the beginning of the RNA chain. The primary pppApApA RNA transcript might be laterally shifted in a 3′ → 5′ direction and the proximal T-residue on the coding DNA strand becomes unpaired. This T-residue can be transcribed and as a consequence four A-residues are obtained at the 5′ end of the transcript. Several rounds of slippage and transcription would result in a 5′ terminal poly(A) sequence. Kassavetis *et al.* further suggested that longer runs of A-residues would probably be generated at increasing ATP concentration. A start of transcription initiaton on the first or second A-residue of the TAAATG-motif of the vaccinia promoter is confirmed by the synthesis of poly(A) head transcripts in the presence of the dinucleotide primers m^7GpppA, GpppA, ApA and GpA (Figure 6B), whereas ApU and m^7GpppG cannot serve as primers. Furthermore, addition of the structural analogue AMP(CH_2)PP does not result in a transcript initiated at the third A-residue migrating at the S1 'start site' position (Figure 6A, lane 1). The gradual decrease in the level of run-off transcription and the simultaneous reduction in the length of the poly(A) head at low concentration of ATP (Figure 5) is compatible with the proposed slippage model. The K_m for ATP incorporation during the poly(A) synthesis (slippage) is presumably higher than that for ATP incorporation during polymerization of the coding part of the ΔC cassette (Figure 5B). Finally, the microheterogeneity in the length of the poly(A) sequences observed *in vitro* and *in vivo* (Schwer *et al.*, 1987) can be attributed to an intrinsic property of the slippage mechanism. Nibbling of the nuclease S1 at the poly(A)/poly(dT) termini of the hybrid, however, cannot be excluded completely.

The incorporation of nucleotides which are not encoded by the genome into mature RNA transcripts is reported to occur in other organisms. cDNA clones of the mitochondrial *cox II* gene of *T.brucei* were shown to contain four U-

residues in the middle of the coding region of the gene which are not encoded by the genome (Benne *et al.*, 1986). The developmentally controlled mitochondrial *CYb* gene of trypanosomes is also subjected to a very unusual process. It was shown that U-residues are incorporated into the untranslated leader of the RNA transcripts which are not encoded by the genome (Feagin *et al.*, 1987).

Additional experiments have to be done to provide conclusive evidence concerning the mechanism of the synthesis of poly(A) head transcripts. The data obtained *in vitro* are consistent with a slippage model, but a novel type of primed initiation mechanism cannot be rule out. Intriguing questions which cannot be explained yet are how an almost discrete length of poly(A) is generated and what its biological function would be.

Materials and methods

Construction of templates

The 11KΔC plasmid was constructed by inserting the *Eco*RI (flushed with T4 DNA polymerase)–*Xba*I fragment containing the vaccinia late promoter of the 11 kd polypeptide (Bertholet *et al.*, 1985) into the *Sma*I–*Xba*I sites of the plasmid p(C2AT)19 (Sawadogo and Roeder, 1985).

The A-less cassette was constructed by insertion of synthetic oligonucleotides with the average composition (CTG) in to the *Eco*RI site (flushed with nuclease S1) of the 11K promoter (*Eco*RI–*Xba*I). The promoter-A-less cassette was inserted into the *Xba*I site of pUC19.

Preparation of whole cell extracts

All extracts were prepared from HeLa S3 cells grown in suspension culture with MEM–Joklik medium containing 5% newborn calf serum. Cells were concentrated, either infected with vaccinia virus (strain WR) at a m.o.i. of 5 or mock-infected and after 1 h diluted with medium to obtain 5×10^5 cells/ml. The incubation was continued for 14 h and whole cell extracts were prepared (Manley and Gefter, 1981). The $(NH_4)_2SO_4$ precipitate was resuspended in 1/20 of the volume of the high-speed supernatant in 25 mM Hepes (pH 7.9), 500 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 5 mM DTT, 17% glycerol and diluted 1:1 in the same buffer without KCl, chromatographed over DEAE–Sephacel (Pharmacia). After concentration of the flow-through by $(NH_4)_2SO_4$ precipitation, the precipitate was resuspended as before and dialysed overnight against 100 vol of the same buffer containing 100 mM KCl. The extracts were aliquoted and stored at −70°C.

In vitro transcription conditions

Standard reactions were performed at 30°C for 45 min in a final volume of 10 μl using 5 μl of dialysed cell extract and a final concentration of 25 mM Hepes (pH 7.9), 50 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM DTT, 8.5% glycerol in the presence of 0.5 mM each of GTP, CTP, ATP and 50 μM UTP, 10 μCi of [α-^{32}P]UTP (800 Ci/mmol) and 100 μg/ml of template DNA. If different incubation conditions were used they are indicated in the figure legend of the respective experiment. Reactions were terminated by addition of 90 μl of 2% sodium dodecylsulphate, phenol extracted and ethanol precipitated in the presence of 0.3 M NaAc (pH 5.2). RNA was precipitated twice, washed with 80% ethanol and dried. Samples were analysed by electrophoresis on a 5% polyacrylamide 7 M urea gel.

S1 mapping

Nucleic acids extracted from a standard reaction mixture were ethanol precipitated in the presence of 0.3 M NaAc and 10 μg of carrier tRNA. S1 mappings were performed according to Maniatis *et al.* (1982). Hybridization was carried out overnight at 44°C and nuclease digestion was at 14°C using nuclease S1 as indicated (Pharmacia). The DNA was degraded prior to S1 mapping by treatment with one unit of RNase free DNase I (Promega) per μg of DNA in a final volume of 100 μl for 15 min at 37°C.

Sources of materials

Radioactive nucleotides were purchased from Amersham, nucleotides and dinucleotides from Sigma, Pharmacia and Boehringer.

Acknowledgements

We would like to thank Riccardo Cortese, Iain Mattaj, Angus Lamond and Lennart Philipson for useful discussions and for critical reading of the manuscript.

References

Beaton,A.R. and Krug,R.M. (1981) *Nucleic Acids Res.*, **9**, 4423–4436.
Benne,R., Van Den Burg,J., Brakenhoff,J.P.J., Sloof,P., Van Boom,J.H. and Tromp,M.C. (1986) *Cell*, **46**, 819–826.
Bertholet,C., Drillien,R. and Wittek,R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 2096–2100.
Betholet,C., Van Meir,E., Heggeler-Bordier,B. and Wittek,R. (1987) *Cell*, **50**, 153–162.
Contreras,R., Cheroutre,H., Degrave,W. and Fiers,W. (1982) *Nucleic Acids Res.*, **10**, 6353–6361.
Feagin,J.E., Jasmer,D.P. and Stuart,K. (1987) *Cell*, **49**, 337–345.
Fogelsong,P.D. (1985) *J. Virol.*, **53**, 822–826.
Furuichi,Y., LaFiandra,A. and Shatkin,A.J. (1977) *Nature*, **266**, 235–239.
Golini,F. and Kates,J.R. (1985) *J. Virol.*, **53**, 205–213.
Hänggi,M., Bannwarth,W. and Stunnenberg,H.G. (1986) *EMBO J.*, **5**, 1071–1076.
Honda,A., Mizumoto,K. and Ishihama,A. (1986) *J. Biol. Chem.*, **261**, 5987–5991.
Kassavetis,G.A., Zentner,P.G. and Geiduschek,E.P. (1986) *J. Biol. Chem.*, **261**, 14256–14265.
Konarska,M.M., Padgett,R.A. and Sharp,P.A. (1984) *Cell*, **38**, 731–736.
Krause,M. and Hirsh,D. (1987) *Cell*, **49**, 753–761.
Makino,S., Stohlman,S.A. and Lai,M.M.C. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4204–4208.
Manley,J.L., Fire,A., Cano,A., Sharp,P.A. and Gefter,M.L. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3855–3859.
Manley,J.L. and Gefter,M.L. (1981) In Chirikjion,G.U. and Papas,T.S. (eds), *Analysis of Nucleic Acids*. Elsevier, New York, pp. 369–382.
Moore,C.L. and Sharp,P.A. (1985) *Cell*, **41**, 845–855.
Moss,B. (1985) In Fields,B.N. (ed.), *Virology*, Raven Press, New York, pp. 685–703.
Murphy,W.W., Watkins,K.P. and Agabian,N. (1986) *Cell*, **47**, 517–525.
Puckett,C. and Moss,B. (1983) *Cell*, **35**, 441–448.
Rohrmann,G. and Moss,B. (1985) *J. Virol.*, **56**, 349–355.
Rosel,J. and Moss,B. (1986) *J. Virol.*, **56**, 830–838.
Sakonju,S., Bogenhagen,D.F. and Brown,D.D. (1980) *Cell*, **19**, 13–25.
Sawadogo,M. and Roeder,R.G. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4394–4398.
Schwer,B., Visca,P., Vos,J.C. and Stunnenberg,H.G. (1987) *Cell*, **50**, 163–169.
Shimotohno,K., Kodama,Y., Hashimoto,J. and Miura,K.I. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 2734–2738.
Slattery,E., Dignam,J.D., Matsui,T. and Roeder,R.G. (1982) *J. Biol. Chem.*, **258**, 5955–5959.
Spaan,W.J.M., Rottier,P.J.M., Horzinek,M.C. and Van der Zeijst,B.A.M. (1982) *J. Virol.*, **42**, 432–439.
Sutton,R.E. and Boothroyd,J.C. (1986) *Cell*, **47**, 527–535.
Venkatesan,S., Baroudy,B.M. and Moss,B. (1981) *Cell*, **125**, 805–813.
Yuen,L. and Moss,B. (1986) *J. Virol.*, **60**, 320–323.

Received on January 5, 1988