Polycistronic vesicular stomatitis virus RNA transcripts

(transcription/polyadenylation/sequence determination)

RONALD C. HERMAN^{†‡}, MANFRED SCHUBERT[†], JACK D. KEENE[§], AND ROBERT A. LAZZARINI[†]

[†]Laboratory of Molecular Biology, National Institute of Neurological and Communicative Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20205; and [§]Department of Microbiology and Immunology, Duke University Medical Center, Durham, North Carolina 27710

Communicated by Robert M. Chanock, May 20, 1980

ABSTRACT A procedure to enrich for the sequences present at the junction between the linked messages in the polycistronic RNAs synthesized *in vitro* by vesicular stomatitis virus (VSV) is described. Analyses of these sequences show that they contain a precise transcript of both the intercistronic dinucleotide and the pentanucleotide 5'-C-U-G-U-U-3', common to the 5'-end of all VSV cistrons, covalently linked to the 3'-side of the intervening poly(A). The data strongly suggest that the VSV transcriptase polyadenylates the mRNAs and can then resume direct and precise transcription of the genome without reinitiation and without skipping nucleotides.

Two different models have been proposed for the transcription of the single-stranded negative-sense genome of vesicular stomatitis virus (VSV). In one model, each of the five capped and polyadenylated messages is separately initiated and terminated; in the other, a large precursor is synthesized and subsequently processed (see ref. 1 for a review). We recently presented electron microscopic evidence for the *in vitro* synthesis of a novel type of polycistronic RNA molecule by purified VSV virions (2). On the basis of those data it was concluded that two or more messages could be covalently linked by an intervening poly(A) sequence. Because the poly(A) is not directly encoded in the VSV genome (3), these linked messages form characteristic loop structures when hybridized to the 42S genomic RNA (Fig. 1).

The potential role of these RNAs as precursors to the monocistronic viral messages has led us to characterize them further by direct biochemical techniques. Our results confirm the earlier observations and show that the linked messages contain faithful transcripts of both the intercistronic region and the pentanucleotide 5'-C-U-G-U-U-3' common to the beginning of each cistron (4) covalently linked to the 3'-side of the intervening poly(A). Taken together, these observations strongly indicate that the VSV transcriptase itself polyadenylates the messages, perhaps by chattering on the seven uridine residues present at the end of each gene, and can then resume direct precise transcription of the genome by simply adding onto the 3'-end of the poly(A).

MATERIALS AND METHODS

***In Vitro* Synthesis of RNA.** *In vitro* transcripts were synthesized in a 0.4-ml reaction mixture containing 0.05 M Tris (pH 8.0), 0.05% Triton N-101, 0.1 M NaCl, 4 mM dithiothreitol, ATP, UTP, and CTP M, 20 μ M GTP, 5 mM MgCl₂, 2 mCi of [α -³²P]GTP (2000-3000 Ci/mmol; 1 Ci = 3.7×10^{10} becquerels), and 200 μ g of the heat-resistant strain of VSV (5); incubation was for 5 hr at 30°C. The RNA was extracted with phenol/chloroform/NaDodSO₄ and precipitated with ethanol. Unincorporated triphosphates were removed by chromatography on Sephadex G-50.

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.

FIG. 1. Electron micrograph of intervening poly(A) loop in hybrid of mRNA and genomic RNA. The 42S VSV genomic RNA was hybridized to polyadenylated *in vitro* product RNA and prepared for electron microscopy as described (2). Because it has no complementary sequence in the genome, the intervening poly(A) forms a single-stranded looped-out region bounded on each side by double-stranded 42S RNA-mRNA duplex. The majority of the single-stranded loops contain 200-500 adenosine residues (2). (Bar represents 0.1 μ m.)

Digestion and Affinity Chromatography. The product RNA was heat-denatured and then digested with RNase T1 (100 units/ml) in 0.01 M Tris, pH 7.5/1 mM EDTA for 30 min at 50°C. The RNase was inactivated with proteinase K (100 μ g/ml) at 37°C for 15 min and then diethylpyrocarbonate (1%) at 37°C for 10 min. The sample was diluted into 0.4 M NaCl/0.01 M Tris, pH 7.4/0.5% NaDodSO₄ and chromatographed on oligo(dT)-cellulose. The bound material was eluted with 0.01 M Tris, pH 7.4/0.05% NaDodSO₄ and precipitated with ethanol. Digestion and chromatography were repeated to further the enrichment.

Rapid RNA Sequence Determination. Fragments isolated after two cycles of RNase digestion and affinity chromatography were digested to completion with nuclease P1 in order to identify the 3'-terminal nucleotide or with RNase T2 to determine its nearest neighbor (6). The mononucleotides were separated by thin-layer chromatography and electrophoresis as described by Keller and Crouch (7). Partial digestion of the RNA with RNase A or formamide was performed as described (8). The fragments were resolved in a 20% polyacrylamide gel (9). Oligonucleotide patterns of randomly cleaved RNA were prepared by using a two-dimensional system consisting of cellulose acetate at pH 3.5 in the first dimension and homochromatography on DEAE-cellulose thin-layer plates in the second (10).

RESULTS

On the basis of electron microscopic analyses, it was estimated that, at most, 10% of the messages synthesized *in vitro* by VSV were linked by intervening poly(A) into polycistronic RNAs (2).

Abbreviation: VSV, vesicular stomatitis virus.

[‡] Present address: New York State Department of Health, Division of Laboratories and Research, Albany, NY 12201.

FIG. 2. Potential sequence across an intercistronic boundary. The lower line shows the nucleotide sequence in the 42S genomic RNA at the junction between the N and NS cistrons (14). The intercistronic dinucleotide is enclosed in the box. The upper sequence is that predicted for a polycistronic RNA with an intervening poly(A) and assumes that the intercistronic dinucleotide is transcribed. The arrowhead indicates the direction of transcription. The sites of RNase T1 cleavage are marked; the locations of the [32 P]phosphates derived from [α - 32 P]GTP are denoted by *.

Therefore, to characterize the sequences contained at the junction of two linked messages, a method had to be devised to enrich for these sequences. The procedure developed is based on sequence data that showed that all five VSV cistrons begin with the common sequence 5'-C-U-G-U-U-3' (4) and end with the common sequence 5'-U-U-U-U-U-U-C-A-U-A-3' (11-13). It is also based on data that showed that the intercistronic regions in the VSV genome contain only the dinucleotide 5'-A-G-3' for the N/NS, M/G, and G/L boundaries or 5'-A-C-3' for the NS/M boundary (14, 15). Thus, a complete transcript across an intercistronic boundary could have the sequence 3'-G-A-C-A-A-U-C-poly(A)-G-U-A-U-5' or 3'-G-A-C-A-A-U-G-poly(A)-G-U-A-U-5' (Fig. 2).

The technique we have used to enrich for these sequences is outlined in Fig. 3. RNA was synthesized *in vitro* by purified heat-resistant VSV virions in the presence of high-specific-activity [α - 32 P]GTP to label the RNA uniformly. Full-length messages linked by poly(A) were detected by electron microscopy among these *in vitro* VSV transcripts at least as frequently as among the Mudd-Summers VSV products (data not shown). After phenol extraction and chromatography on Sephadex G-50, the RNA was heat-denatured and digested with RNase T1. Because of the specificity of this enzyme, the proposed sequence (Fig. 2) should be cleaved to yield the fragment 3'-

FIG. 3. Experimental design. RNA was synthesized *in vitro* by the heat-resistant strain of VSV in the presence of [α - 32 P]GTP. The product RNA was extracted twice with phenol/chloroform/NaDodSO₄ and then chromatographed on Sephadex G-50 to remove the unincorporated triphosphates. The RNA was then heat-denatured and digested with RNase T1 prior to chromatography on oligo(dT)-cellulose. Digestion and selection were repeated a second time to enrich for the intercistronic sequences.

Table 1. Proportion of RNA bound to oligo(dT)-cellulose

Exp.	Total	Bound*	% of total	Polycistronic RNA,† %
1	7×10^7	12,900	0.018	5
2	7×10^8	58,950	0.009	3
3	5×10^8	52,200	0.010	3

* cpm eluted from oligo(dT)-cellulose after second cycle of digestion and affinity chromatography.

† Calculated by assuming that the average VSV transcript contains approximately 300 guanosine residues.

pGpA-C-A-A-U-C-poly(A)-5' [or 3'-pGppoly(A)-5'] uniquely labeled in only one place with [32 P]phosphate (indicated by *). The polyadenylated RNA was selected by chromatography on oligo(dT)-cellulose and then redigested with RNase T1 and reselected on oligo(dT)-cellulose to further the enrichment. The proportion of the radioactivity retained by the oligo(dT)-cellulose after the second cycle of digestion and selection is shown in Table 1 for three separate experiments. It is estimated from these data that perhaps 3-5% of the messages synthesized in these three experiments were linked by poly(A). This frequency is not significantly different from the previous estimate that up to 10% of the messages were linked by poly(A) (2) and implies that the selected sequences were not derived from very small artifactual transcription products. Electrophoresis of the labeled RNase T1 oligonucleotides containing poly(A) in a 20% gel indicated that they were at least 70 and probably up to 200 nucleotides long (data not shown).

A portion of the fragments selected by the oligo(dT)-cellulose chromatography was completely digested with nuclease P1 in order to determine the origin of the [32 P]phosphate. Thin-layer chromatography/electrophoresis (7) of the digestion products showed that the labeled phosphate was contained solely in 5'-GMP. This indicates that the [32 P]phosphate was not randomized during the *in vitro* synthesis reaction and that the [α - 32 P]GTP did not contain detectable levels of other labeled triphosphates. Complete digestion of a second portion with RNase T2 and analysis of the products by thin-layer chromatography/electrophoresis showed that the [32 P]phosphate was transferred uniquely from the 5'-side of the guanosine to 3'-AMP. These results demonstrate that all of the fragments must terminate with 3'-pGpA-5'.

Partial digestion of the RNA with pancreatic RNase A yielded two fragments (Fig. 4). The sizes of these fragments were determined by comparison to those generated by random, partial hydrolysis with formamide. The locations of the sites cleaved by RNase A indicate that pyrimidines are present at positions 3 and 6 from the 3'-end of the sequence. The interpretation of the sequence thus far obtained is shown at the right in Fig. 4. The locations of the purines and pyrimidines are entirely consistent with those in the predicted sequence 3'-G-A-C-A-A-U-C-5'.

A further determination of the sequence was made by using a two-dimensional system consisting of electrophoresis on cellulose acetate at pH 3.5 in the first dimension and homochromatography on DEAE-cellulose plates in the second (10). The RNA, partially hydrolyzed in carbonate buffer at pH 9, was resolved in the first dimension according to base composition and in the second according to size. Because the addition of each nucleotide to the sequence results in a characteristic shift to the left or the right in the first dimension depending on the base, the sequence can be directly read from the autoradiogram (Fig. 5). Each spot was subsequently eluted from the plate and its identity was confirmed by DEAE-cellulose paper electrophoresis at both pH 1.7 and pH 3.5 (data not shown). The sequence obtained from these data, 3'-G-A-C-A-A-U-C-A-A-5', is

FIG. 4. Partial RNase digestion of polyadenylylated fragments. The RNA fragments selected as outlined in Fig. 3 were partially digested with RNase A and the resulting oligonucleotides were resolved in a 20% polyacrylamide gel as described (9) (lane 1). The polyadenylylated fragments were also randomly cleaved with formamide (8) to generate a "ladder" (lane 2). The sequence analysis is shown at the right; N, a purine; Y, a pyrimidine; *, [^{32}P]phosphate. The locations of the tracking dyes xylene cyanol (XC) and bromophenol blue (BPB) are indicated at the far right.

precisely that predicted for a complete transcript across an intercistronic boundary and includes a transcript of the intercistronic dinucleotide. Because the RNA was selected, after complete RNase T1 digestion, by its ability to bind repeatedly to oligo(dT)-cellulose, it must have contained a large tract of adenosine residues of which only the last three were resolved in the two-dimensional system. Indeed, the uniform spacing of the larger fragments (above band 6) in the formamide ladder (Fig. 4) are consistent with the presence of a substantial tract of adenosine residues at the 5' end of the RNA.

DISCUSSION

The results presented here directly show that some *in vitro* VSV transcription products contain a precise transcript of both the intercistronic dinucleotide and the common pentanucleotide

FIG. 5. Sequence analysis in two dimensions. The polyadenylylated fragments were partially hydrolyzed in carbonate buffer at pH 9 for 15 min at 90°C. The resulting oligonucleotides were resolved in the first dimension on cellulose acetate at pH 3.5 and in the second on a DEAE-cellulose thin-layer plate (10). The sequence is shown at the right.

covalently attached to the 3'-side of an intervening poly(A). The sequence detected at the junction of the linked messages corresponds to that predicted for the N/NS, M/G, and G/L intercistronic boundaries. However, we did not find unambiguous evidence for the fragment 3'-G-poly(A)-5' predicted from the sequence for the NS/M junction of the San Juan isolate even though loops were frequently observed at this position in the electron microscope. The reason for this is not known but may reflect sequence differences between the Mudd-Summers isolate of VSV (Indiana) (16) used for the electron microscopy, the heat-resistant isolate of VSV (Indiana) (5) used here, and the San Juan isolate of VSV (Indiana) used for the sequence experiments (15).

The synthesis of linked messages *in vitro* and *in vivo* (ref 2; unpublished observations) as well as of *in vivo* L message linked to the 59-base extracistronic oligonucleotide (unpublished observations) could indicate that a large precursor is synthesized and processed. Indeed, evidence already exists to show that transcription across the leader/N junction can occur *in vitro* (ref. 17; unpublished observations). Clearly, these observations are only suggestive of a processing mechanism. Nevertheless, they provide additional insight into the mechanism of VSV transcription. The data suggest that the VSV transcriptase, itself, polyadenylylates the RNA, most probably without leaving the template. Seven uridine residues are found in the template at the end of each cistron, and it has been proposed that poly-

adenylation of the nascent transcripts occurs by "chattering" on these residues (13, 14). In fact, such chattering must also occur during leader RNA synthesis because four adenosine residues are occasionally found at the 3'-end of this transcript even though only three uridines are present in the template (18-20).

We cannot unequivocally rule out an alternative mechanism in which some other enzyme polyadenylates the nascent VSV mRNA and that VSV polymerase is able to use the 3'-end of the poly(A) tail to prime further transcription. However, evidence for such a model is completely lacking. Cordycepin triphosphate inhibits VSV transcription and polyadenylation equally, and viral extracts will not adenylate exogenous polynucleotides *in vitro* (21). Direct attempts to prime VSV transcription with poly(A)⁺ viral messages have also been unsuccessful (unpublished observations). Furthermore, the electron microscopic analyses show that only transcripts of adjacent genes are linked by poly(A) (2), suggesting that the polymerase remains associated with both the mRNA and the template during polyadenylation and resumes direct transcription without reinitiation and without skipping nucleotides to yield a precise transcript across the intercistronic region. If, on the other hand, the poly(A) were added to the message by another enzyme, then the transcriptase would have to reattach to the end of the newly adenylated RNA and resume direct transcription of the genome at *precisely* the place where it had stopped.

The extent of polyadenylation appears to be markedly affected by inhibitors of methylation such as S-adenosylhomocysteine (22). This same compound also increases the frequency at which linked transcripts are found even though the poly(A) can be up to 20 times longer than normal (unpublished observations). Thus, the mechanisms by which the poly(A) and the linked transcripts are synthesized appear to be intimately related but additional experimentation will be required to understand this relationship.

Although the results presented here are suggestive of a processing mechanism, they also may be explained by illegitimate or aberrant transcriptive events that are not directly related to the synthesis of viral mRNA. A final decision on the relevance of these linked messages to the mechanism of mRNA synthesis may depend upon a future demonstration that the polycistronic

transcripts can be processed into mature, capped, and methylated viral mRNAs.

J.D.K. was supported by grants from the National Institute of Allergy and Infectious Diseases (AI 16099) and from North Carolina United Community Services.

1. Banerjee, A. K., Abraham, G. & Colonno, R. J. (1977) *J. Gen. Virol.* **34**, 1-8.
2. Herman, R. C., Adler, S., Lazzarini, R. A., Colonno, R. J., Banerjee, A. K. & Westphal, H. (1978) *Cell* **15**, 587-596.
3. Marshall, S. & Gillespie, D. (1972) *Nature (London) New Biol.* **240**, 43-45.
4. Rhodes, D. P. & Banerjee, A. K. (1976) *J. Virol.* **17**, 33-42.
5. Petric, M. & Prevec, L. (1970) *Virology* **41**, 615-630.
6. Schubert, M., Keene, J. D., Lazzarini, R. A. & Emerson, S. U. (1978) *Cell* **15**, 103-112.
7. Keller, W. & Crouch, R. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 3360-3364.
8. Schubert, M., Keene, J. D. & Lazzarini, R. A. (1979) *Cell* **18**, 749-757.
9. Donis-Keller, H., Maxam, A. M. & Gilbert, W. (1977) *Nucleic Acids Res.* **4**, 2527-2538.
10. Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965) *J. Mol. Biol.* **13**, 373-398.
11. McGeoch, D. J. & Turnbull, N. T. (1978) *Nucleic Acids Res.* **5**, 4007-4024.
12. Rose, J. K. & Iverson, L. (1979) *J. Virol.* **32**, 404-411.
13. Schubert, M., Keene, J. D., Herman, R. C. & Lazzarini, R. A. (1980) *J. Virol.* **34**, 550-559.
14. McGeoch, D. J. (1979) *Cell* **17**, 673-681.
15. Rose, J. K. (1980) *Cell* **19**, 415-421.
16. Lazzarini, R. A., Weber, G. H., Johnson, L. D. & Stamminger, G. M. (1975) *J. Mol. Biol.* **97**, 289-307.
17. Testa, D. & Banerjee, A. K. (1978) *Biochem. Biophys. Res. Commun.* **82**, 655-664.
18. Colonno, R. J. & Banerjee, A. K. (1978) *Cell* **15**, 93-101.
19. Rowlands, D. J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4793-4797.
20. Keene, J. D., Schubert, M. & Lazzarini, R. A. (1980) *J. Virol.* **33**, 789-794.
21. Murphy, M. F. & Lazzarini, R. A. (1974) *Cell* **3**, 77-84.
22. Rose, J. K., Lodish, H. F. & Brock, M. L. (1977) *J. Virol.* **21**, 683-693.