Volume 15 Number 10 1987 Nucleic Acids Research

A procedure for selective full length cDNA cloning of specific RNA species

Anita Schmid, Roberto Cattaneo and Martin A.Billeter*

Institut für Molekularbiologie I, Universität Zürich, Hönggerberg, CH-8093 Zürich, Switzerland

Received March 10, 1987; Revised and Accepted April 23, 1987

ABSTRACT

A method allowing routine establishment of full length and functionally competent cDNA clones of particular mRNAs from small preparations of polyadenylated RNA is described. Pairs of synthetic primers are used for first and second strand synthesis. They include sequences complementary to the 3′ terminal regions of the mRNAs and of the full length first cDNA strands, respectively and bear a few additional nucleotides at their 5′ ends. After synthesis of both cDNA strands in one tube, they are precisely trimmed back with T4 DNA polymerase in presence of only two nucleoside triphosphates, to yield sticky ends fitting into a vector plasmid cleaved with two restriction endonucleases. The procedure was first applied to the simultaneous cloning of all five major measles virus (MV) mRNA species from a persistently infected cell line. Two thirds of all clones contained full length MV-specific cDNAs. Screening of less than 200 clones was sufficient to obtain several independent clones corresponding to each mRNA, except for gene F which was represented only once.

INTRODUCTION

For a variety of applications it is highly desirable to derive full length and functionally competent cDNA clones, corresponding to specific mRNA species present at low abundance, from small samples of polyadenylated RNAs. Full length cDNAs could greatly facilitate the structural and functional characterization of variant mRNAs which occur in a variety of human diseases as consequence of mutations, resulting in defective translation and/or non-functional translation products. Our immediate interest in an efficient method of cloning variant mRNAs stems from the study of subacute sclerosing panencephalitis (SSPE). This is a rare and generally fatal disease caused by MV persistence in the human brain several years after the acute infection. It is thought that combinations of particular mutations in defective MVs are instrumental for the establishment and/or maintenance of the disease (1-5). To obtain an understanding of the mutations and their biological consequences in SSPE-related defective MV strains, it is necessary to clone, analyse and functionally express all the variant mRNAs. However, the application of unspecific cDNA cloning procedures for this purpose would be prohibitively laborious. We therefore devised a specific cloning procedure based on particular synthetic oligonucleotides for priming of first and second strand synthesis. This procedure turned out to be both efficient and very specific, yielding a large number of plasmids containing full length cDNA inserts. We started with 5 μg of poly A-selected RNA from a cell line persistently infected with a

SSPE-derived strain of MV. The cell line contained an average of 1000 copies per cell of each of the MV-specific mRNAs, roughly one tenth of the amount present in lytically infected cells. Seventy-five clones were obtained when one twentieth of the cDNA recovered from the 5 μg of poly A-selected RNA was used for ligation and transfection. Two thirds of the clones contained inserts corresponding to all full length MV mRNA species aimed at, which range in length from 1470 to 2372 nucleotides. In addition, a clone containing a 4.6 kb cDNA insert derived from a rare polycistronic MV transcript was identified.

MATERIALS AND METHODS

Source of RNA and RNA preparation

Total cellular RNA was obtained from cell line IP-3-Ca (5) as described previously (6). The RNA was treated with RNase-free DNase and proteinase K (4) and poly A-selected (7) by one passage over a column of oligo-dT cellulose (type 3, Collaborative Res., Lexington, Mass., USA). The amount of MV-specific RNAs present in this cell line was determined by quantitative Northern blot analysis with internal RNA standards (4).

Oligonucleotide primers

Specific oligonucleotide primers for five MV mRNAs (table 1) were prepared with an automated synthesizer (Applied Biosystems), purified by polyacrylamide gel electrophoresis, treated with pronase (Calbiochem) in presence of 0.5% sodium dodecyl sulfate, extracted with phenol and precipitated with ethanol. The forward primers were complementary to 19-22 nucleotides of the 3′ terminal region adjacent to the polyadenylate stretch of the MV transcripts, and included the additional four nucleotides CCGG at their 5′ end (table 1). The reverse primers corresponded to the first 21-23 nucleotides at the 5′ end of the MV transcripts and contained the additional nucleotides CGTT at their 5′ end. The oligonucleotides were phosphorylated using polynucleotide kinase and either γ^{32}P-ATP or unlabelled ATP (7).

Synthesis of first and second cDNA strands

2.5 μg of polyadenylated RNA were mixed with 2.5 pmol of each forward primer (a 500 fold molar excess on average) in 70 μl of 10 mM Hepes-HCl pH 6.9, 0.2 mM EDTA (HE). The mixture was heated for 2 min to 90° and then cooled in ice water. cDNA synthesis was performed after adjusting the reaction mixture to contain 50 mM Tris-HCl pH 7.5, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 0.5 mM each of dATP, dCTP, dGTP and dTTP, 125 units of RNAsin (Biotec, Madison, USA) and 600 units of cloned MLV reverse transcriptase (BRL, Bethesda, USA). The final volume was 150 μl; incubation was for 5 min at 25°, 5 min at 30° and then 90 min at 37°. KOH was then added to a concentration of 75 mM and the mixture was heated for 60 min to 65° in order to hydrolyze the RNA. After neutralization with HCl and Hepes-HCl pH 6.9, 2.5 pmol of each reverse primer were added. The preparation was then heated for 15 min to 50° and slowly cooled to room temperature over a period of one hour. The mixture was adjusted to contain 100mM Hepes-HCl pH 6.9, 10 mM $MgCl_2$, 2.5 mM fresh DTT and 0.5 mM each of

Table 1 First and second strand primers used for cDNA cloning.

Gene	Forward primers	Reverse primers
N	5´**CCGG**TTATAACAATGATGGAGGG	5´**CGTTA**GGATTCAAGATCCTATTATC
P	5´**CCGG**TTATAATGGATTTAGGTTG	5´**CGTTA**GGAACCAGGTCCACACAG
M	5´**CCGG**TTGTTTAGTTTGCCGGGGAG	5´**CGTTA**GGAGCAAAGTGATTGCCTC
F	5´**CCGG**TTAATTAACTACCGATATTGT	5´**CGTTA**GGGCCAAGGAACATACAC
H	5´**CCGG**TTAATTCTGATGTCTATTTCAC	5´**CGTTA**GGGTGCAAGATCATCCAC

Bold letters indicate nucleotides not complementary to MV mRNAs. A and T residues adjacent to the 5´ terminal G and C residues are underlined. The sequences of the ends of the MV genes are listed in (4). The genes are (in this genomic order) N: nucleocapsid (1693 nucl.); P: phosphoprotein (1655 nucl.); M: matrix (1470 nucl.); F: fusion (2372 nucl.); H: hemagglutinin (1959 nucl.).

fresh four nucleoside triphosphates in a total volume of 350 µl. Twenty-five units of the Klenow fragment of E. coli DNA polymerase I were added and incubation proceeded at 15° overnight. The reaction was stopped by adding EDTA to 15 mM; after addition of 2 µg of carrier RNA and phenol extraction, the cDNAs were precipitated with ethanol.

Size selection of cDNAs and T4 DNA polymerase reaction

The cDNAs were dissolved in 90 µl of 10 mM Tris-HCl pH 8.0, 0.1mM EDTA, and 10 µl of 330 mM Tris acetate pH 7.9, 670 mM potassium acetate (10x T4 salt solution) were added. The preparation was chromatographed on a column of Sephacryl S 500 (Pharmacia, Uppsala) prepared in a 1 ml Pasteur pipette. (The column had been pretreated with 10 µg of sonicated salmon sperm DNA and thoroughly rinsed with T4 salt solution). The Cherenkov radiation in each of the 70 µl fractions was determined; aliquots from pools of 3 fractions each were analyzed by polyacrylamide gel electrophoresis. Pools containing mostly cDNAs larger than 1 kb were combined (total volume 420 µl) and adjusted to 10 mM magnesium acetate, 0.5 mM DTT, 0.1 mg/ml autoclaved gelatine and 0.1 mM dATP and dTTP. Four units of T4 DNA polymerase (Pharmacia, Uppsala) were added and after 15 min at 20° the preparation was deproteinized by incubation for 10 min at 37° with 50 µg/ml proteinase K in 20 mM EDTA, 0.5% sodium dodecyl sulfate. After addition of 0.5 µg of carrier RNA, phenol extraction, ethanol precipitation and washing twice with 80% ethanol, the cDNAs were dissolved in HE.

Vector preparation, ligation, transfection and characterization of cloned inserts

Vector Bluescript M13 (plus orientation, see fig. 1 for a summary of the properties of the vector) was obtained from Stratagene (San Diego, California). It was cleaved successively with the restriction endonucleases AccI and XcyI, treated with phosphatase (7) and the small AccI-XcyI fragment was eliminated by chromatography on Sephacryl S 500. Ten ng (5 fmol) of this vector DNA preparation were ligated with varying amounts of the cDNAs (see results for optimal ratios)

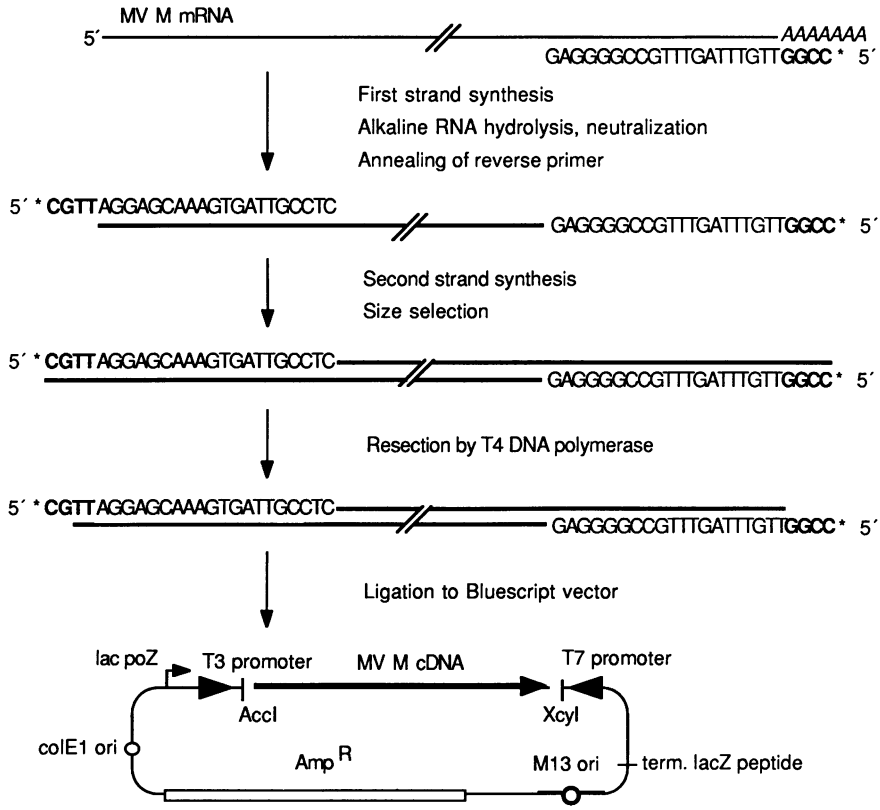

Figure 1 Overview of the cloning procedure.

using standard protocols (7). The ligated material was transfected into competent JM 109 cells (8). The screening for inserts was carried out either by restriction endonuclease digestion of small preparations of plasmid DNA (9) or by colony hybridization (7), using either the labelled primer oligonucleotides or specific RNA probes (4). The sequences surrounding the vector-insert junctions were examined by the chain termination method (10) directly on alkali-treated plasmid DNA (11) using M13 sequencing primers (8); for the analysis of internal sequences of the cDNA inserts specifically synthesized primers were used.

RESULTS

Primed first and second cDNA strand synthesis

An overview of our cDNA cloning procedure is given in fig. 1. It was our aim to obtain full length cDNA clones of all 5 major MV mRNA species from a minimal amount of poly A-selected cellular RNA. Thus, first cDNA strand synthesis was primed with a mixture of synthetic

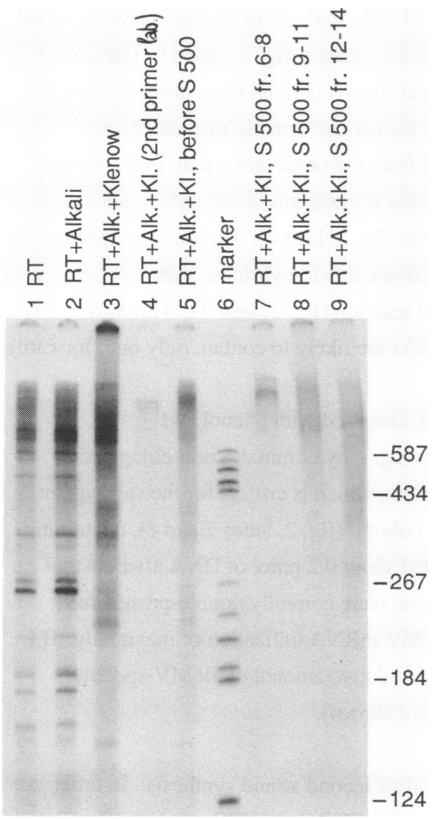

Figure 2 Polyacrylamide gel electrophoresis of cDNA products. The products of a reaction containing only two pairs of primers are shown. This protocol was used for efficient cloning of the intercistronic boundary of a long bicistronic transcript encompassing sequences of genes P and M. The forward primers were primer M forw. (Table 1) and a primer hybridizing within the M coding region (741-719 bases downstream of the P-M intergenic nucleotides). The reverse primers were primer P rev. and a primer corresponding to the P coding region (196 to 175 bases upstream of the P-M intergenic nucleotides). About 10% of the "labelled first strand" reaction was used for lanes 1-3, about 10% of the "labelled second strand" reaction for lane 4, about 2% of the pooled reactions for lane 5 and about 7% of the pooled fractions of the S 500 column for lanes 7-9. Lengths of the molecular weight markers shown in lane 6 are indicated on the right of the gel. Resolution was on a 5% polyacrylamide, 7M urea denaturing gel.

oligonucleotides (forward primers, table 1). Two reaction mixtures were set up and run in parallel to monitor both cDNA synthesis steps independently. For the first strand synthesis one reaction mixture contained forward primers radioactively labelled at their 5´ends, whereas in the other mixture forward primers phosphorylated with nonradioactive ATP were used. The primers, at large molar excess, were first annealed with 2.5 µg each of poly A-selected RNA from a cell line persistently infected with a SSPE-derived defective MV strain. To minimize losses, the ensuing steps (first strand synthesis, hydrolysis of template RNA by alkali, neutralization, hybridization with a mixture of reverse primers and second strand synthesis using the Klenow fragment of E. coli DNA polymerase I) were all carried out in the original reaction tubes without any transfer or intermediate purification procedure. In the reaction mixture containing non-labelled forward primers, ^{32}P-labelled reverse primers were used for second strand synthesis. Small samples of the preparations at different reaction steps were analyzed by polyacrylamide gel electrophoresis. Fig. 2 gives an illustration of such an analysis. (The results shown there are from an experiment in which only two pairs of primers had been employed, 4; all other results shown are from the experiment involving the five pairs of primers shown in table 1).

First strand synthesis reaction yielded some long cDNA products in the size range expected , but in addition also large amounts of shorter products accumulated (fig. 2, lane 1). The short products likely were a mixture of incompletely elongated strands derived from primers correctly hybridized to MV mRNA, and of strands derived from primers incorrectly associated and elongated on cellular RNA species. This was deduced from the observation that, in the second strand reaction, the elongation of the newly added reverse primers amounted only to about one twentieth of that monitored during first strand synthesis (fig. 2, lane 4). However, it appears that a large fraction of the short first strand reaction products not able to hybridize with reverse primers nevertheless increased in size during the second strand reaction (fig. 2, lane 3). This was probably due to self-primed synthesis; thus, these cDNAs are likely to contain only one (forward) primer.

At this point both reaction mixtures were pooled, extracted with phenol and chromatographed on Sephacryl S 500. This gel chromatography eliminates non-elongated primers, some short cDNAs and nucleoside triphosphates, which is critical for the subsequent liberation of sticky ends. Fractions 6-11 of the S 500 column (fig. 2, lanes 7 and 8), containing long duplex DNA molecules were combined and yielded about 0.2 pmol of DNA after ethanol precipitation. Only about 5% of this material (0.01 pmol) were correctly double-primed full length products. This corresponds to an efficiency of MV mRNA utilization of maximally 10% because, from quantitative Northern blot analysis the cumulative amount of all MV-specific mRNAs had been estimated as about 0.1 pmol (data not shown).

Precise resection of DNA duplex molecules

Correct cDNAs presumably possess flush ends after second strand synthesis. In order to liberate 5′ overhanging "sticky ends" complementary to the CCGG- and CG- overhanging ends of the linearized vector plasmid, the products were treated with T4 DNA polymerase. The presence of both dATP and dTTP at high concentrations limited the 3′ exonucleolytic strand degradation effectively at the position of the first A or T residue. To ensure that even after incidental exonucleolytic removal of more than one residue beyond that point the synthesizing activity of T4 DNA polymerase could restore the DNA strands, all primers were constructed such that at least two A or T residues followed the terminal C and G residues (underlined in table 1); this is the reason why the reverse primers all contained two extra T residues not present at the 5′ termini of the MV mRNAs. The proper incubation conditions for the T4 polymerase reaction were tested on homopolymer regions from cloned cDNA inserts (results not shown).

Ligation with vector and transfection

To optimize ligation of the tailored cDNAs into the cloning vector prepared by double digestion with XcyI and AccI, two sets of different conditions were examined: use of cleaved vector with or without phosphatase treatment prior to ligation, and variation of the ratio of cDNA to vector. These results are summarized in table 2. Phosphatase treatment of the vector (lines 6-9) almost completely abolished the high background of blue colonies due to plasmid without insert,

Table 2 Transformation efficiency of JM 109 cells with Bluescript vector and cDNA inserts.

Number of colonies	blue	white
1 Vector (0.5 fmol), closed circular, +ligase	1500	0
2 Vector (5 fmol), cleaved[a], -ligase	5	0
3 Vector (5 fmol), cleaved[a], +ligase	790	1
4 Vector (5 fmol), cl.[a], +cDNA (8 fmol[b]), +ligase	1200	75
5 Vector (5 fmol), cl.[a], +cDNA (40 fmol[b]), +ligase	800	110
6 Vector (5 fmol), cl.[a], dephosphorylated, -ligase	1	0
7 Vector (5 fmol), cl.[a], dephosphorylated, +ligase	3	0
8 Vector (5 fmol), cl.[a], dephos., +cDNA (8 fmol[b]), +lig.	1	60
9 Vector (5 fmol), cl.[a], dephos., +cDNA (40 fmol[b]), +lig.	5	46

[a] The vector was cleaved in the cloning site region by AccI and XcyI and the resulting small fragment was removed.
[b] The amount of cDNA was calculated from the combined labelled forward and reverse primers recovered after the S 500 column. Since only about 5% of the cDNAs contained reverse primers, only about 0.4 and 2 fmol of correct cDNAs are estimated in the total 8 and 40 fmol, respectively.

probably derived from low amounts of vector not fully cleaved with both restriction endonucleases. In contrast, the yield of white colonies due to plasmids with insert was decreased by a factor of less than two by use of dephosphorylated vector. The preferable ratio of DNAs for ligation was a slight molar excess of total insert to vector (lines 4 and 8). Although under these conditions the ratio of correct insert to vector was estimated to be less than 1 : 10 considering the relatively low yield of correct second strand synthesis, an increase of this ratio did not result in a comparable increase of colony numbers (lines 5 and 9). This was most likely due to relatively short cDNA products containing only one ligatable "sticky end" (derived from forward primers) which were ligated at one vector end, thus competing with correct cDNA molecules. In further experiments only phosphatase-treated vector and relatively low ratios of inserts to vector were used.

Characterization of cloned inserts

Thirty-six white colonies were characterized in a preliminary screening by restriction analysis of plasmid DNA isolated by a quick lysis procedure. Twenty-two of these indicated a full length MV mRNA insert; eighteen corresponded to MV P, two to N and two to H mRNA inserts. Five plasmids showed inserts longer than 1500 base pairs, but their restriction fragment patterns were not uniquely compatible with MV- type inserts. Eight plasmids showed only short inserts of 200 to 1000 base pairs, and one an unexpectedly large insert of 4.6 kb. To identify also inserts corresponding to genes M and F, colony hybridization was performed with a total of 191 clones using specific probes. One clone showed hybridization to the F, four to the M probe. All of the nine clones examined by sequence analysis after the preliminary characterization contained

the expected cDNA corresponding to one of the five entire MV mRNAs. The 4.6 kb long insert encompassed the entire N-, P- and the majority of the M-mRNA sequence, including the intergenic boundaries, and was apparently formed on a tricistronic MV transcript onto which the P primer (fully represented in the insert) had hybridized incorrectly to a GC- rich untranslated region, about 170 nucleotides upstream the 3′ end of the M gene sequence.

DISCUSSION

We have used pairs of specific primers to generate full length cDNA copies of MV mRNAs. In an elegant approach reported recently by Scharf et al. (12), similar pairs of primers containing target sequences for restriction endonucleases have been used to preferentially amplify short segments of human genomic DNA by repeated cycles of denaturation, primer annealing and extension by DNA polymerase I. In principle, it should be possible to amplify DNA copies obtained from rare mRNAs by combining our procedure with the one described by Scharf et al. (12). However, for our present application this proved not to be necessary. A possible difficulty in combining the two approaches could arise by failure to attain complete denaturation of DNAs, which in our case were about ten times longer than the DNA segments aimed for in the amplification procedure.

In order to produce sticky ends flanking the cDNAs for ligation into the vector, we tailored our blunt-ended cDNAs using T4 DNA polymerase. This has some advantages over the use of primers containing "external" restriction sites (12). First, no care has to be taken to avoid in the primer constructs any restriction enzyme target sites which are present in the insert sequence. Second, in our primers not more than four non-fitting 5′ terminal nucleotides are necessary, thus reducing nonspecific primer hybridization. Third, the high specificity of our method (65% specific inserts) is in all probability primarily due to the selection against molecules not possessing two fitting "sticky" ends. When reverse primers hybridize nonspecifically to first cDNA strands such that extended stretches of their 5′ ends do not form a duplex with the first cDNA strand, T4 polymerase resection will not liberate a precisely fitting "sticky" CG-end.

It should be pointed out that our procedure is not dependent upon using forward primers complementary to the 3′ terminal region of mRNA or reverse primers corresponding to the 5′ termini of mRNA. In order to efficiently clone an intercistronic region of a MV bicistronic transcript (4), we used in addition to terminal primers, a pair of primers complementary to internal sequences of RNA and first cDNA strands, respectively (see legend to fig. 2 for details). In this experiment we obtained all possible combinations of inserts bracketed either by terminal or internal primers. These results suggest a) that the internal primers were able to hybridize efficiently to the RNA and the first cDNA strand, respectively, and b) that first strands which had not been completely elongated on the bicistronic transcript were efficiently resected by T4 DNA polymerase to the point where the internal reverse primer had hybridized. The possibility to use primers fitting to internal sequences would be convenient in several possible applications of the

method, such as the insertion of a precisely delimitated sequence in an expression vector, e.g. for the expression of eukaryotic coding sequences in prokaryotic promoter-ribosome binding site contexts. Since the method allows cloning of rarely occurring long transcripts, it should also prove useful for unambiguous definition of suspected long unspliced, partially or alternatively spliced transcripts.

Our experiments, using mixtures of primers, show that the efficiency of cloning varies considerably for different mRNA species. According to the relative abundance of the MV mRNA species in the cell line studied herein, the highest yield of inserts was expected for gene N sequences, whereas all other cDNAs should have been cloned with three to five times lower efficiencies. However, almost 80% of the MV specific clones corresponded to the P gene, whereas the other genes together shared the remaining 20%. Obviously several factors influence the yield of every particular cDNA: the efficiency of primer hybridization, the ease of primer elongation through particular sequences and the length of the RNA regions to be cloned. However, RNA length does not appear to be a decisive factor, since the number of full length clones obtained does not correlate inversely with the mRNA lengths given in the legend to table 1 and since inserts derived from rare long RNA transcripts were obtained.

So far we have not observed that sequences between the primer hybridization points were missing in the cloned inserts, a frequent occurrence in our hands when applying a frequently used cloning protocol (13). However, in some cases as mentioned above for the 4.6 kb insert, we did obtain clones derived from forward primers hybridized to RNA regions with only partial complementarity.

To date we have only used two particular overhanging ends, but in principle it should be possible to employ any pair of ends, provided they consist only of two different nucleotides. They should be followed, in the primer sequence, by at least two consecutive other nucleotides.

Finally, it should be mentioned that the MV inserts in the cloning vector are readily transcribed with T3 RNA polymerase in presence of unmethylated "cap" nucleoside triphosphate (G(5′)ppp(5′)G). These transcripts produced, in a reticulocyte- derived in vitro translation system, immunoprecipitable proteins with a similar molar efficiency as MV- specific mRNA extracted from MV-infected cells (G. Rebmann and K. Baczko, unpublished results).

ACKNOWLEDGMENTS

We thank Dr. Stephen Udem for the total RNA of cell line IP-3-Ca, Gabi Rebmann for the quantitative determination of the MV-specific RNAs of this line, Dr. Horst Hornig for the synthesis of specific oligonucleotides, Dr. Knut Baczko for communicating unpublished results, Dr. Markus Aebi and Dr. Charles Samuel for critical reading of the manuscript and Fritz Ochsenbein for the photograph. This work was supported by grant 3.141-085 of the Schweizerische Nationalfonds and by the Kanton Zürich.

*To whom correspondence should be addressed

REFERENCES

1. ter Meulen, V., Stephenson, J.R. and Kreth, H.W. (1983) Compr. Virol. **18**, 105-185.
2. Baczko, K., Liebert, U.G., Billeter, M.A., Cattaneo, R., Budka, H. and ter Meulen, V. (1986) J. Virol. **59**, 472-478.
3. Cattaneo, R., Schmid, A., Rebmann, G., Baczko, K., ter Meulen, V., Bellini, W.J., Rozenblatt, S. and Billeter, M.A. (1986) Virology **154**, 97-107.
4. Cattaneo, R., Rebmann, G., Schmid, A., Baczko, K., ter Meulen, V. and Billeter, M.A. (1987) EMBO J., 681-688.
5. Sheppard, R.D., Raine, C.S., Bornstein, M.B., and Udem, S.A. (1986) Proc. Natl. Acad Sci. USA **83**, 7913-7917.
6. Udem, S.A. and Cook, K.A. (1984) J. Virol. **49**, 57-65.
7. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, N. Y.
8. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103-119.
9. Holmes, D. S. and Quigley, M. (1981) Anal. Biochem. **114**, 193-197.
10. Sanger, F., Coulsen, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. **143**, 161-178.
11. Chen, E.J. and Seeburg, P.H. (1985) DNA **4**, 165-170.
12. Scharf, S.J., Horn, G.T. and Erlich, H.A. (1986) Science **233**, 1076-1078.
13. Gubler, U. and Hoffmann, B.J. (1983) Gene **40**, 1-8.