VIROLOGY 223, 376–380 (1996)
ARTICLE NO. 0490

SHORT COMMUNICATION

ELKE MÜHLBERGER,[1] SABINE TROMMER, CHRISTA FUNKE, VIKTOR VOLCHKOV, HANS-DIETER KLENK, and STEPHAN BECKER

Institut für Virologie, Philipps-Universitat Marburg, Robert-Koch-Str. 17, 35037 Marburg, Germany

Received April 18, 1996; accepted July 9, 1996

The 3′ and 5′ ends of Marburg virus (MBG)-specific mRNA species have been determined using reverse transcription-PCR, rapid amplification of cDNA ends, or the reverse ligation-mediated PCR procedure after removal of cap structures with tobacco acid pyrophosphatase. The polyadenylation sites of all MBG-specific mRNAs were strictly conserved and corresponded to the predicted transcriptional stop signals of genomic RNA. Determination of the 5′ ends of the mRNA species showed that mRNA synthesis started precisely at the first nucleotide of a highly conserved transcriptional start site. The 5′ ends of the mRNA species can build a stable secondary structure with the conserved nucleotides always located in the stem region of a hairpin. Nucleotide substitutions in the conserved 5′ regions are accompanied by compensatory mutations of the complementary nucleotide thus leading to a conservation of the secondary structures. Compensatory mutations were also found when 5′ ends of mRNA of MBG strain Musoke were compared with MBG strain Popp or the closely related Ebola virus, indicating that the secondary structures will be conserved even if the sequence is altered.

Marburg virus (MBG) is the prototype of the filovirus family, which belongs to the order Mononegavirales. MBG causes a severe hemorrhagic disease in monkeys and humans with high fatality rates (*1*). Common features of the order Mononegavirales are a nonsegmented negative-stranded (NNS) RNA genome which is encapsidated by a nucleoprotein, a viral envelope composed of a lipid bilayer with inserted surface proteins, and an RNA-dependent RNA polymerase responsible for transcription and replication of the viral genome (*2*). In contrast to rhabdo- and paramyxoviruses, which are known to possess three nucleocapsid proteins, the filoviruses contain an additional protein associated with the nucleocapsid (*3*). The four nucleocapsid proteins of MBG are the nucleoprotein (NP; *4*), the viral proteins VP35 and VP30, and the L protein (*5*). In addition to these, three structural proteins were detected: a single surface protein (GP; *6*) and the putative matrix proteins VP40 and VP24. The genomic RNA of MBG is 19,105 nucleotides in length (for reference see EMBL Nucleotide Sequence Database, Accession No. Z12132) and is transcribed into seven monocistronic mRNA species (*7*) encoding the viral proteins mentioned above. Sequence analysis of the genomic RNA revealed that the genes are separated by short intergenic regions with exception of the overlapping genes encoding VP30 and VP24. At the putative gene boundaries highly conserved regions are located which might serve as transcription start and stop signals (*7*). These signals were identified by comparison with the previously determined sequence of the 3′ and 5′ ends of NP mRNA (*8*). However, it was noted that the 5′ end of NP mRNA was preceded by 2 unidentified nucleotides. In comparison with the genome-encoded putative transcriptional start signals of the other genes these 2 preceding nucleotides were not conserved. Since data concerning transcription start and stop signals of MBG were obtained mainly by sequencing of genomic RNA, it was necessary to complement them by a detailed analysis of isolated mRNA. The present study was carried out to prove (i) whether the putative transcriptional signals correlate with the 3′ and 5′ ends of MBG mRNA and (ii) whether nonconserved nucleotides are located at the extreme 5′ termini of the mRNA species. This paper contains the experimental determination of the exact 3′ and 5′ termini of all mRNA species of MBG and a computer-assisted analysis of the sequencing data.

Protocols for the exact determination of RNA termini have recently been developed (*9, 10*) and were used for the experiments described in this paper. For this purpose, MBG-infected E6 cells were lysed 5 days p.i., and total RNA was isolated using the protocol described by Sanchez and Kiley (*11*). Since the titer of MBG stocks were rather low (2×10^5 PFU/ml), it first had to be shown that the yield of MBG-specific mRNA was sufficient for the planned analyses. For detection

[1] To whom correspondence and reprint requests should be addressed. E-mail: muehlber@mailer.uni-marburg.de.

0042-6822/96 $18.00

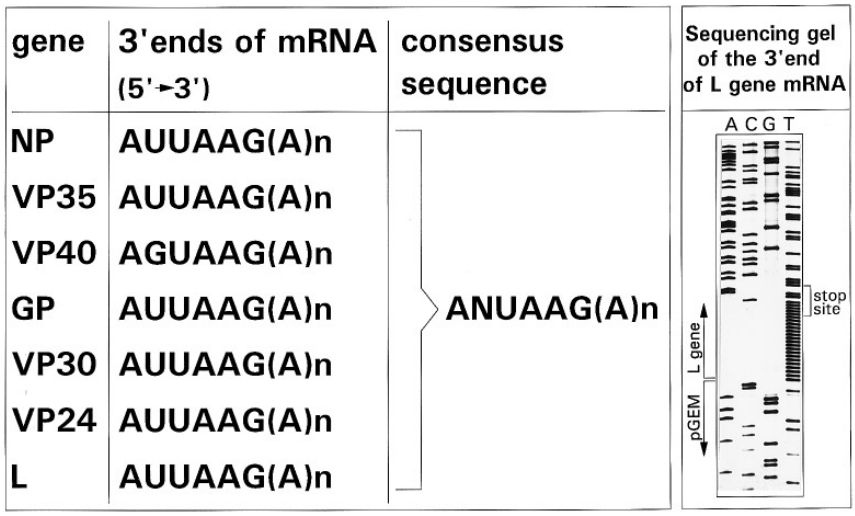

FIG. 1. Analysis of 3′ ends of MBG-specific mRNA. 1.5×10^8 E6 cells (a cloned Vero cell line, ATCC CRL 1586) were infected with MBG strain Musoke at an m.o.i. of 0.1 PFU per cell. Five days p.i. cells were lysed and RNA was isolated as described by Sanchez and Kiley (*11*). RT-PCR was performed using a commercial RNA-PCR kit (Perkin–Elmer). Four micrograms of total RNA was used for reverse transcription which was performed in a volume of 20 μl at 42° for 40 min using an oligo(dT) primer (21 dT residues; final concentration 2.5 μM) supplied with a *Bam*HI site. After cDNA synthesis samples were heated to 99° for 7 min, cooled on ice, mixed with PCR buffer, $MgCl_2$ (final concentration 2 m*M*), and the different MBG gene-specific primers supplied with a *Bam*HI site (final concentration 0.5–1 μM). The volume was adjusted to 99.5 μl, and the samples were heated for 5 min at 95°. Amplification was carried out by 35 cycles of 94° for 1 min, 40° for 1 min, and 72° for 2 min after addition of 0.5 μl Ampli*Taq* DNA polymerase (2.5 U). The amplified fragments (160–360 nucleotides in length) were digested with *Bam*HI and cloned into the *Bam*HI site of the vector pGEM3Zf(+). For each gene several clones were sequenced. Amplifying of the 3′ end of the L mRNA required a nested PCR. The L gene-specific primer used for the first PCR was located 353 nucleotides upstream of the deduced 3′ end, the primer for the second PCR bound 95 nucleotides upstream of the deduced 3′ end. As second primer for the nested PCR the oligo(dT) primer mentioned above was used. Both amplification reactions were performed as described above. The determined sequences are depicted in mRNA sense. At the right of the table a 6% denaturing polyacrylamide gel shows the cloned 3′ end of the L-specific mRNA in vRNA sense.

of MBG mRNA species, Northern blot analyses were performed with MBG gene-specific DNA probes (genes coding for NP, VP35, and VP24) coupled with digoxygenin (DIG DNA labeling kit; Boehringer Mannheim). Using the probes specific for the first and second gene of MBG (NP and VP35), it was possible to detect the corresponding mRNA species, whereas the sixth gene of MBG (VP24) was not detectable by Northern blot analysis (data not shown). For other NNS RNA viruses it is known that mRNA transcription starts only at a single promoter site, resulting in a decreasing relative abundance of mRNA from 3′ to 5′ end of genomic RNA (*12*). Our Northern blot analyses indicated therefore that MBG genes located at the 3′ end of the genome (e.g., NP) are more frequently transcribed compared to the genes located at the 5′ end (e.g., VP24 and L).

To overcome the sensitivity problem, a reverse transcription-PCR (RT-PCR) was performed amplifying fragments of NP and VP24 mRNA. Obtained PCR products were analyzed by the Southern blot technique. Both NP- and VP24-specific DNA fragments could be detected by this method (data not shown). These results suggested that sufficient MBG mRNA was produced for analyzing the 3′ ends by RT-PCR. E6 cells were infected with MBG strain Musoke and lysed 5 days p.i. The total RNA was isolated and used for analyzing the 3′ ends of MBG-specific mRNA species by RT-PCR as described for Fig. 1. With exception of the L gene, the 3′ ends of all mRNA species could be amplified by one PCR reaction. Amplifying of the 3′ end of L mRNA required a nested PCR. Thus, it was possible to identify the MBG L mRNA that had not been detected before. In Fig. 1 the results are summarized. The polyadenylation sites of all genes were strictly conserved and corresponded to the predicted transcription stop signals $(UA/CAUUC(U)_5$; *7*). Interestingly, a single nucleotide substitution in the conserved region of the VP40 gene (U → G) was found in the sequence of genome (*7*) and mRNA. Only the five adenosine residues of the poly(A) tail following the conserved region are encoded by the genome, indicating that polyadenylation was performed by stuttering of the viral polymerase as reported for other NNS RNA viruses (*13*).

Analysis of the 5′ ends of mRNA was performed using either the reverse ligation-mediated PCR procedure after removal of cap structures with tobacco acid pyrophosphatase (TAP-RLPCR; *10*) or a rapid amplification of cDNA ends (5′ RACE technique). An advantage of the TAP-RLPCR method is that ligation of a known RNA linker sequence to the 5′ end of mRNA occurs at the RNA and not at the cDNA level. Thus, possible errors during reverse transcription of the extreme 5′ end of RNA are avoided. In addition removal of the 5′ cap leads to a higher accuracy in determination of the first nucleotide. For TAP-RLPCR, E6 cells were infected with MBG strain

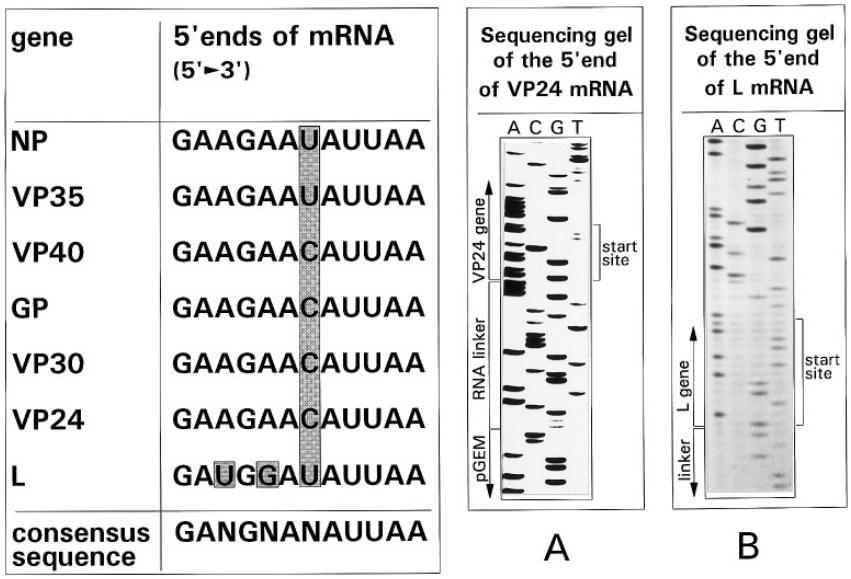

FIG. 2. Analysis of 5′ ends of MBG-specific mRNA. For TAP-RLPCR (*10*) total RNA was purified from MBG-infected E6 cells as described for Fig. 1. Ten micrograms of total RNA was treated with 1 U tobacco acid pyrophosphatase (TAP; Biozym) in TAP reaction buffer (50 m*M* sodium acetate, pH 6.0, 1 m*M* EDTA, 0.1% mercaptoethanol, 0.01% Triton X-100, 2 m*M* ATP) in a volume of 10 μl for 30 min at 37°. One microliter of this sample was ligated with 100 ng of an RNA linker (5′-GGGCAUAGGCUGACCCUCGCUGAAA) in a total volume of 10 μl using 3 U of T4 RNA ligase (Boehringer Mannheim) in RNA ligase buffer (50 m*M* Tris–HCl, pH 7.5, 10 m*M* $MgCl_2$, 20 m*M* DTT, 0.1 mg/ml BSA, 2 U/μl RNasin). After an incubation period of 16 hr at 16° 10% of the sample was used for RT-PCR which was performed using a commercial RNA PCR kit (Perkin–Elmer). Reverse transcription was done in a volume of 20 μl for 35 min at 52° with MBG gene-specific primers binding about 300 nucleotides downstream of the putative 5′ ends of the mRNA species (final concentration 0.75 μM). Thereafter samples were heated to 99° for 7 min, cooled on ice, mixed with PCR buffer, $MgCl_2$ (final concentration 2 m*M*), and a primer complementary to the RNA linker carrying a *Bam*HI site (final concentration 0.15 μM). The volume was adjusted to 99.5 μl, and the samples were heated for 5 min at 95°. 0.5 μl Ampli*Taq* DNA polymerase (2.5 U) was added, and PCR was run under conditions described below. Following the first amplification a nested PCR was performed with other gene-specific oligonucleotides supplied with a *Bam*HI site creating DNA fragments of 124–230 bp in length. Both PCR were carried out under the following conditions: 35 cycles of 1 min at 94°, 1 min at 64–67°, 2 min at 72°. The amplified fragments were digested with *Bam*HI and cloned into the vector pGEM3Zf(+). The 5′ ends of GP, VP30, and L mRNA were analyzed using the 5′-Amplifinder RACE kit (Clontech) according to the supplier's instructions. Briefly, 10 μg of total RNA was incubated with a GP-, VP30-, or L-specific primer binding approximately 300 nucleotides downstream of the putative 5′ ends (final concentration 1 μM) for 10 min at 65°. Thereafter, cDNA synthesis was performed for 30 min at 52°. After purification of cDNA a single-stranded DNA linker was ligated to the different cDNA species for 16 hr at 22° using T4 RNA ligase. One microliter of a $\frac{1}{10}$ dilution of the ligation sample was used for the following PCR, which was run for 35 cycles at 94° for 1 min, 60° for 1 min, and at 72° for 2 min using MBG gene-specific primers carrying a *Bam*HI site and a primer complementary to the single-stranded DNA linker (final concentration 0.15 μM). The generated fragments, which were approximately 200 nucleotides in length, were digested with *Bam*HI and *Eco*RI, and cloned into the vector pGEM3Zf(+). The sequence of the inserts representing the 5′ termini of the mRNA was determined by enzymatic sequencing. The analyzed sequences are depicted in mRNA sense. Variable nucleotides are marked by a gray bar. (A) 6% denaturing sequencing gel showing the 5′ end of VP24-specific mRNA determined by TAP-RLPCR (mRNA sense); (B) 6% denaturing sequencing gel showing the 5′ end of L-specific mRNA determined by the 5′ RACE method (mRNA sense).

Musoke and lysed 5 days p.i. The total RNA was isolated and used for determination of the 5′ ends of NP, VP35, VP40, and VP24 mRNA species by TAP-RLPCR as described for Fig. 2. The 5′ ends of GP, VP30, and L mRNA were analyzed using the 5′-Amplifinder RACE kit (Clontech) according to the supplier's instruction. With both methods it was possible to determine the 5′ ends of all MBG mRNA species. It is shown in Fig. 2 that the sequence of the 5′ ends of six mRNA species was highly conserved. However, position 7 of the sequence was represented by either a uridine or a cytosine residue (Fig. 2, bar). Only the 5′ end of the L mRNA showed two additional nucleotide exchanges at positions 3 and 5 (Fig. 2, marked by gray boxes). In a previous study, the 5′ end of NP mRNA has been analyzed using the primer extension method in combination with chemical sequencing (*8*). The following transcription start site was found: 5′-NNGAAGAAUAUUAA (mRNA sense) consisting of the conserved region GAAGAAYAUUAA and two preceding nucleotides which could not be identified. In contrast to these results, the present data show that mRNA synthesis of all analyzed genes starts precisely at the first conserved nucleotide of the predicted start sites obtained by sequencing of genomic RNA (Fig. 2; *7*). Neither with the 5′ RACE nor with the TAP-RLPCR method could any additional nucleotide at the 5′ end of the conserved region be detected.

Computer-assisted predictions based on the 5′ ends revealed secondary structures which are similar for all MBG-specific mRNA species (Fig. 3A). The conserved nucleotides of the 5′ ends are almost completely located in the stem region of a hairpin. As mentioned above (Fig.

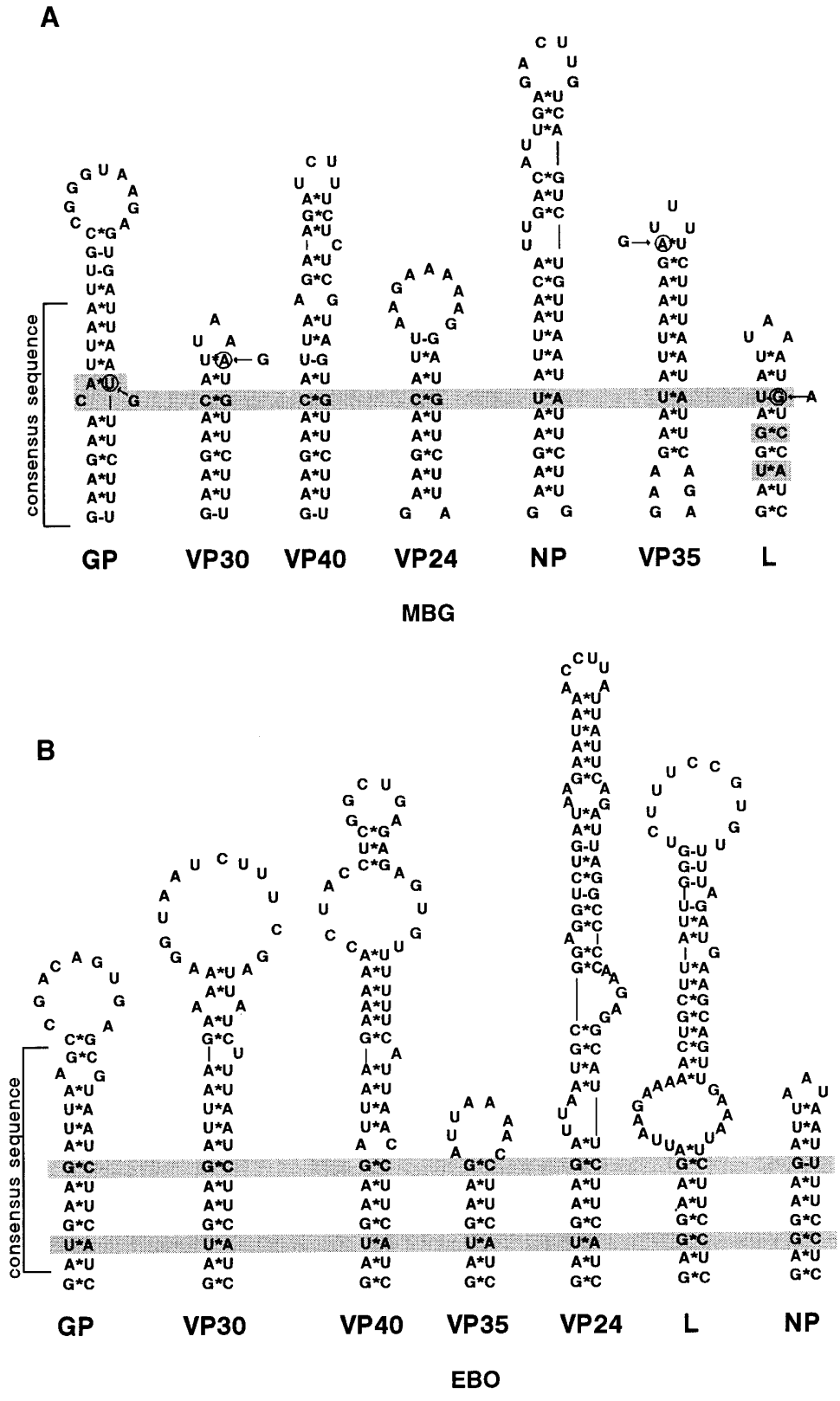

FIG. 3. Predicted secondary structure of the 5′ ends of MBG and EBO mRNA species. (A) The nucleotide sequence of the 5′ ends of mRNA species of MBG strain Musoke was analyzed with Heidelberg Unix Sequence Analysis Resources, release 3.0, program FOLDRNA. Nucleotides surrounded by circles are exchanged in MBG strain Popp. Exchanged nucleotides are indicated by arrows. (B) The predicted 5′ termini of EBO mRNAs were analyzed according to Volchkov *et al.* (*17*) and Sanchez *et al.* (*15*) except that the first two nucleotides were omitted. Base pairing is marked by asterisks, G-U pairing is indicated by lines. The positions of the conserved regions are marked by a bracket at the left side of each scheme. Compensatory mutations in the conserved regions are marked by bars.

2), in six of the seven mRNA species only one nucleotide of the conserved 5′ end is variable (position 7). Interestingly, the alteration of the respective nucleotide is always accompanied by a compensatory mutation of the complementary nucleotide in the stem–loop element. In the case of VP30, VP40, and VP24 a C-G base pairing is found, whereas NP, VP35, and L showed at this place a U-A or a U-G base pair (Fig. 3A, bar). One exception is GP mRNA because its cytosine residue at position 7 is not involved in base pairing. In this context the 5′ end of the L mRNA is of special interest. Although the sequence of the conserved region is altered at positions 3 and 5, the stem structure is retained by simultaneous mutations of the base pairing nucleotides (Fig. 3, marked by gray boxes). When the determined sequences of MBG strain Musoke were compared to the closely related MBG strain Popp (*14;* EMBL Nucleotide Sequence Database, Accession No. Z29337), three nucleotide exchanges could be found (Fig. 3A, surrounded by circles). These nucleotide exchanges did not alter the complementarity in the stem regions thus leading to an almost identical secondary structure formation in both strains. The most interesting exchange is the U to G transversion in the GP gene, because in Popp strain this transversion led to a C-G base pair at position 7, whereas the adenosine residue at position 8 is not involved in base pairing. As mentioned above in Musoke strain the situation is inverted: nucleotide 7 is not involved in base pairing but nucleotide 8 is (Fig. 3A). These results indicate that the secondary structure of the 5′ ends is conserved even when the nucleotide sequence is altered, suggesting an important role of these stem–loop elements for the reproduction of MBG. The other member of the family of filoviruses, Ebola virus (EBO), also showed a highly conserved nucleotide sequence at the predicted 5′ ends of its mRNA species (5′-GAG/UGAAGAUUAA) which is similar but not identical to MBG (*15*). Thus, there are two nucleotide exchanges at positions 3 and 7 (underlined). Computer-based analysis of the deduced 5′ ends also suggested the formation of stem–loop structures resembling those found with MBG (Fig. 3B). As it is shown in Fig. 3, all mutations in the stem regions of MBG and EBO mRNAs are always compensatory mutations (marked by bars), thus maintaining a stable base pairing. These data underline that such secondary structures might serve as important regulatory elements for filoviruses. So far no data are available elucidating the function of these elements during translation initiation and/or transcription. For picornaviruses it is reported that secondary structures of the 5′ ends of their RNA play an important role in initiation of the translation process (*16*).

ACKNOWLEDGMENTS

The authors thank Angelika Lander for her expert technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 286) and by the Alexander von Humboldt-Stiftung through a fellowship to V. Volchkov.

REFERENCES

1. Martini, G. A., and Siegert, R., Eds., "Marburg Virus Disease." Springer, New York, 1971.
2. Pringle, C. R., *Arch. Virol.* 117, 137–140 (1991).
3. Elliott, L., Kiley, M. P., and McCormick, J. B., *Virology* 147, 169–176 (1985).
4. Becker, S., Huppertz, S., Klenk, H.-D., and Feldmann, H., *J. Gen. Virol.* 75, 809–818 (1994).
5. Mühlberger, E., Sanchez, A., Randolf, A., Will, C., Klenk, H.-D., and Feldmann, H., *Virology* 187, 534–547 (1992).
6. Will, C., Mühlberger, E., Linder, D., Slenczka, W., Klenk, H.-D., and Feldmann, H., *J. Virol.* 67, 1203–1210 (1993).
7. Feldmann, H., Mühlberger, E., Randolf, A., Will, C., Kiley, M. P., Sanchez, A., and Klenk, H.-D., *Virus Res.* 24, 1–19 (1992).
8. Sanchez, A., Kiley, M. P., Klenk, H.-D., and Feldmann, H., *J. Gen. Virol.* 73, 347–357 (1992).
9. Edwards, J. B. D. M., Delort, J., and Mallet, J., *Nucleic Acids Res.* 19, 5227–5232 (1991).
10. Fromont-Racine, M., Bertrand, E., Pictet, R., and Grange, T., *Nucleic Acids Res.* 21, 1683–1684 (1993).
11. Sanchez, A., and Kiley, M. P., *Virology* 157, 414–420 (1987).
12. Glazier, K., Raghow, R., and Kingsbury, D. W., *J. Virol.* 21, 863–871 (1977).
13. Banerjee, A. K., *Microbiol. Rev.* 51, 66–87 (1987).
14. Bukreyev, A. A., Volchkov, V. E., Blinov, V. M., Dryga, S. A., and Netesov, S. V., *Arch. Virol.* 140, 1589–1600 (1995).
15. Sanchez, A., Kiley, M. P., Holloway, B. P., and Auperin, D. D., *Virus Res.* 29, 215–240 (1993).
16. Jang, S. K., Pestova, T. V., Hellen, C. U., Witherell, G. W., and Wimmer, E., *Enzyme* 44, 292–309 (1990).
17. Volchkov, V. E., Blinov, V. M., Kotov, A. N., Chepurnov, A. A., and Netesov, S. V., IXth International Congress of Virology, Glasgow, P52-2, 1993.