Volume 322, number 1, 41–46 FEBS 12389 May 1993

The VP35 and VP40 proteins of filoviruses

Homology between Marburg and Ebola viruses*

A.A. Bukreyev, V.E. Volchkov, V.M. Blinov and S.V. Netesov

Institute of Molecular Biology, 633159 Koltsovo, Novosibirsk Region, Russian Federation

Received 9 March 1993

The fragments of genomic RNA sequences of Marburg (MBG) and Ebola (EBO) viruses are reported. These fragments were found to encode the VP35 and VP40 proteins. The canonic sequences were revealed before and after each open reading frame. It is suggested that these sequences are mRNA extremities and at the same time the regulatory elements for mRNA transcription. Homology between the MBG and EBO proteins was discovered.

Marburg virus; Ebola virus; Filoviridae; Nucleotide sequence; Amino acid sequence

1. INTRODUCTION

Marburg (MBG) and Ebola (EBO) viruses belong to the Filoviridae family [1]. These viruses cause similar hemorrhagic fevers with a high mortality rate [2].

Both viruses have seven structural proteins. GP is the only glycosylated protein [3,4]. The gene order determined for MBG by Feldmann et al. [5] is the following: 3′-NP-VP35-VP40-GP-VP30-VP24-L-5′. The proteins are synthesized from mRNAs transcribed from the negative-strand RNA genome [3,4].

The nucleotide sequence of the MBG Musoke strain NP gene was reported by Sanchez et al. [6] and of the L gene by Muhlberger et al. [7]; for the EBO Mayinga strain the sequence of the NP gene was published by Sanchez et al. in [8]. In our previous work we reported the nucleotide sequence of the EBO Mayinga strain 3′ end of the GP gene [9]. In this work, we present the nucleotide sequences of MBG and EBO genomic RNA fragments, which encode the VP35 and VP40 proteins, as well as deduced amino acid sequences and the homology between these proteins of both viruses.

Correspondence address: A.A. Bukreyev, Institute of Molecular Biology, Koltsovo, 633159, Novosibirsk Region, Russian Federation. Fax: (7) (3832) 328 831.

*The presented sequences of the fragments of MBG and EBO genomes were published in the EMBL Data Library (X64406, 1992, and X61274, 1991, respectively).

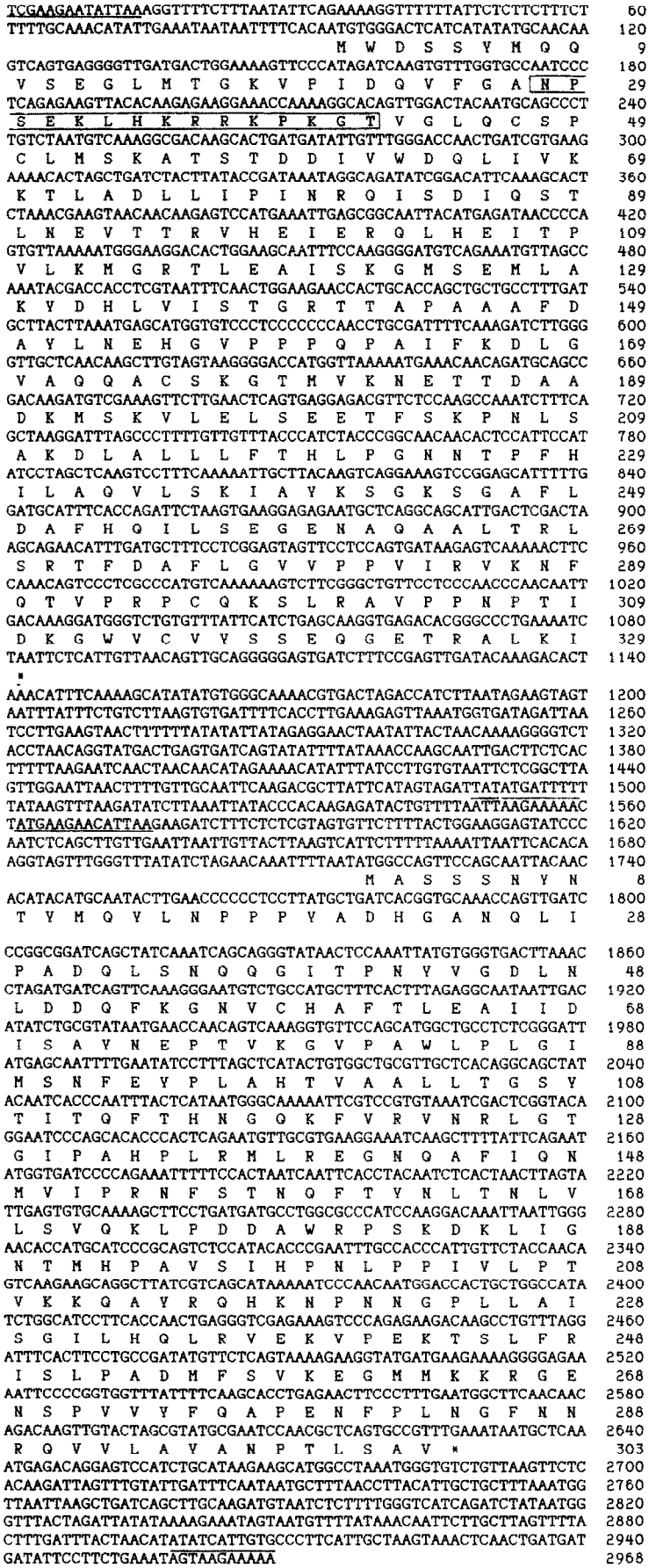
2. MATERIALS AND METHODS

The Popp strain of MBG and the Mayinga strain of EBO were received from the Belarus Institute of Epidemiology and Microbiology (Minsk, Belarus). Purification of the virus, isolation of the genomic RNA, synthesis, cloning and sequencing of cDNA were carried out as in [10] for MBG and as in [9] for EBO.

3. RESULTS AND DISCUSSION

The primary structures of the MBG and EBO genomic RNAs were determined by sequencing partly overlapping cDNA-containing recombinant plasmids. We found seven long open reading frames (ORFs) in the cDNA of both viruses which corresponded to the seven known viral proteins.

The fragment of the cDNA MBG genomic RNA sequence with ORF2 and ORF3 encoding the VP35 and VP40 proteins is shown in Fig. 1. The context of the initiation AUG codon for ORF1 is not the most favourable for translation initiation for eukaryotic ribosomes; the AUG codon for ORF2 corresponds to the Kozak rule [11]. The length of the putative VP35 polypeptide is 329 amino acids (aa) and that of VP40 is 303 aa. The calculated mol.wt. are 36,149 and 33,734 Da, respectively, which approximates the results of SDS-PAGE [5]. Computer analysis of the full-length cDNA sequence determined canonic regions 3′-CU$^{U}_{A}$C$^{U}_{C}$U$^{G}_{A}$-UAAUU-5′ and 3′-U$^{A}_{C}$AUUCUUUUU-5′ (the sequences are given for negative-strand RNA) before and after each ORF, respectively. The former of the above sequences constitutes a part of transcriptional start signal 3′-NNCUNCNUNUAAUU-5′, described in [5] for MBG Musoke strain and shown to be mRNA extremi-



ties. The latter corresponds to the transcriptional stop signal and mRNAs extremity sequence 3′-UAAUUCUUUUU-5′ [5] (Fig. 1). We suggested that the conserved sequences from Fig. 1 are mRNA extremities and signals for initiation and termination of transcription simultaneously.

The sequence of the cDNA EBO genomic RNA fragment is shown in Fig. 2. The VP35 gene ORF starts from the AUG codon located at positions 100–102 (calculated mol. wt. of the putative protein is 37,362 Da). Despite the fact that the first AUG codon corresponds better to the consensus sequence for eukaryotic initiation [11] than the second one located at 157–159, we assume that the VP35 synthesis starts from the latter. In this case the length of the putative polypeptide is 321 aa; the calculated mol. wt. of 35,277 Da fits better that evaluated by SDS-PAGE analysis [4]. The comparison of the putative polypeptide with the corresponding MBG polypeptide (see below; Fig. 3) supports this assumption. The AUG codon near the 5′ terminus of the VP40 ORF corresponds to the consensus sequence for eukaryotic ribosomes [11]. The length of the putative VP40 polypeptide is 326 aa; the calculated mol. wt. of 35,182 Da is less than that estimated by SDS-PAGE (40,000 Da) [4]. A conceivable reason for this difference may be post-translation modification of synthesized protein.

As with MBG, computer analysis of the full-length cDNA of EBO genomic RNA revealed the canonic sequence 3′-$^{A}_{G}$CU$^{A}_{C}$CUUCUAAUU-5′ before all the seven ORFs and the sequence 3′-UAAUUCUUUUU-5′ after 6 of the 7 ORFs (with the exception of the 7th ORF) (data not shown). Since these sequences are similar to the transcription start signal, 3′-UACUCCUUCUAAUU-5′ and to the stop signal, 3′-UAAUUCUUUUU-5′, respectively, which were described for the NP gene of EBO and shown to be NP mRNA extremity sequences [8], we suggest that the canonic sequences from Fig. 2 are transcriptional start and stop signals and mRNA extremities for the VP35 and VP40 mRNAs simultaneously. The VP35 and VP40 genes of EBO share a common interesting trait: the stop signal of VP35 and the start signal of VP40 are overlapping (Fig. 2). A similar feature was found with the MBG VP30 and VP24 genes ([5]; Bukreyev et al., unpublished data).

Both pairs of the genes have long 3′ and 5′ untranslated regions. Table I shows the lengths of these regions.

Significant homology between both polypeptides of the two viruses was revealed. The alignments between the MBG and EBO proteins are shown in Fig. 3. The amino acid homology between the VP35 proteins makes up 33% (the putative N-end of the EBO VP35 corresponds to the second methionine of the ORF); homology between the VP40 proteins is 27%. The comparison of hydropathic plots [12] of the amino acid sequences is presented in Fig. 4. It is obvious from the figure that there is similarity between the MBG and EBO proteins. The plot of the MBG VP35 shows a striking hydrophilic domain (28–42) (Fig. 1); the analogous domain in EBO is located at 57–76 (Fig. 2). Since no amino acid homology between these two regions was revealed, the existence of the domains may be due to the functional similarity of these regions in both viruses.

The position of the MBG and EBO VP35 and VP40 genes in the genome, and the fact that the VP40 constitutes 39% of total virion protein in MBG [3–5,13], support the assumption that VP35 is a protein analogous to the P proteins of paramyxoviruses, and VP40 is the main matrix protein [3,5,8,13]. However, although considerable homology was revealed earlier between the L proteins of MBG, EBO and paramyxoviruses [7,14,15], we failed to discover any homology of the MBG and EBO VP35 and VP40 proteins with the protein sequences of paramyxoviruses.

The significant homology between the MBG and EBO VP35 and VP40 proteins confirms the evolutionary relationship between these two members of the Filoviridae family.

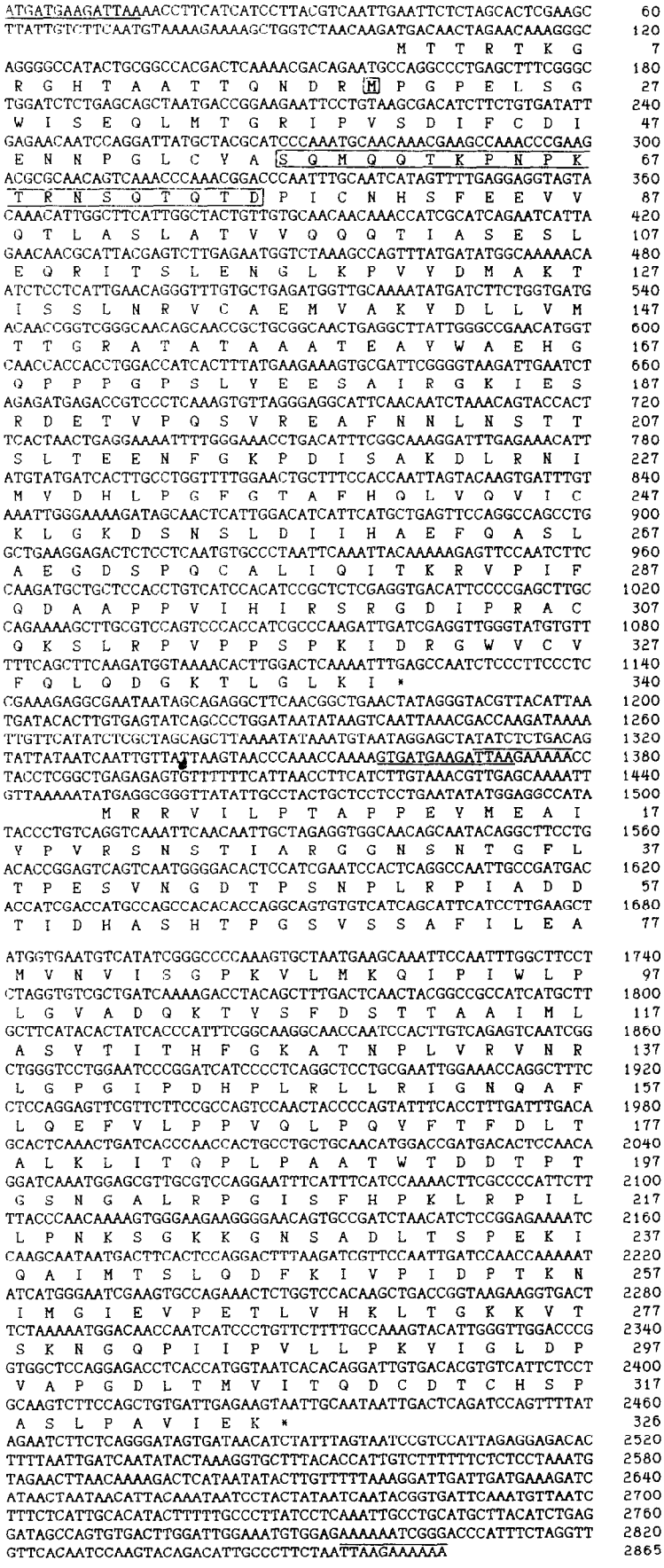
Table I

The lengths of the 3′ and 5′ non-coding regions of the VP35 and VP40 genes of MBG and EBO

	3′ end	5′ end
MBG VP35	93	479
EBO VP35	156 (99)	259
MBG VP40	155	343
EBO VP40	91	438

←

Fig. 1. The fragment of the MBG genome (cDNA plus-strand) with the predicted amino acid sequences of the VP35 and VP40 proteins (shown beneath the nucleotide sequence). The putative start signals for the RNA-dependent RNA-polymerase are underlined and the stop signals are overlined. The striking hydrophobic region is boxed.



```
MBG  ------------------MWDSSVMQQVSEGLMTGKVPIDQVFGANPSEKLHKRRKPKG  41
EBO  MTTRTKGRGHTAATTQNDR.PGPELSGWI..Q. ..RI VSDI.CDIENNPGLCYASQMQ  60

MBG  TVGLQCSPCLMSKATSTDDIVWDQLIVKKTLADLLIPINRQISDIQSTLNEVTTRVHEIE 101
EBO  QTKPNPKTRNSQTQ.D-PICNHSFEE.VQ...S.ATVVQQ TIASE.LEQRI.S------ 113

MBG  RQLHEITPVLKMGRTLEAISKGMSEMLAKYDHLVISTGRTTAPAAAFDAYLNEHGVPPPQ 161
EBO  -LENGLK..YD.AK.ISSLNRVCA..V....L..MT...A..T...TE..WA..Q...G 172

MBG  PAIFKDLGVAQQACSKGTMVKNETTDAADKMSKVLELSEETFSKPNLSAKDLALLLFTHL 221
EBO  .SLYEESAIRGKIE.RDET.PQSVRE.FNNLNSTTS.T..N G..DI.....RNIMVD.. 232

MBG  PGNNTPFHILAQVLSKIAVKSGKSGAFLDAFHQI-LSEGENAQAALTRLSRTFDAFLGVV 280
EBO  ..FG.A .Q.V. IC.LG- DSN.LDIIH.EF.AS.A..DSP.C..IQITKRVPI.QDAA 291

MBG  PPVIRVKNFQTVPRPCQKSLRAVPPNPTIDKGWVCVYSSEQGETRALKI             329
EBO  ....HIRSRGDI..A.... . P...S.K..R.....FQLQD.K LG...          340
(A)

MBG  MASS---SNYNTYMQYLNPPPYADHGA---NQLIPADQLSNQQGITPNYVGDLNLDD---  51
EBO   RRVILPTAPPE..EAIV.VRSNSTI.RGG SNTGFLTPESVN.D..SNPLRPIA..TID  60

MBG  ---QFKGNVCHAFTLEAIIDISAVNEPTVKGVPAWLPLGIMSNFEYPLAHTVAALLTGSV 108
EBO  HASHTP.S.SS..I. .MVNVISGPKVLM.QI.I... .VADQKT.SFDS.T..IMLA.. 120

MBG  TITQFTHNGQKFVRVNRLGTGIPAHPLRMLREGNQAFIQNMVIPRNFSTNQFTVNLTNLV 168
EBO  ...H.GKATNPL.......P...D....L..I.....L.EF.L.PVQLPQV..FD..A.K 180

MBG  LSVQKLPDDAWRPSKDKLIGNTMHPAVSIHPNLPPIVLP--TVKKQAVRQHKNPNNGPLL 226
EBO  .IT.P..AAT.TDDTPTGSNGALR.GI.F..K.R..L..NKSG..GNSADLTS.EK---- 236

MBG  AISGILHQLRVEKVPEKTSLFRISLPADMFSVKEGMMKKRGENSPVVYFQAPENFPLNGF 286
EBO  -.QA.MTS.QDF.IVPIDPTKN.MGIEVPETLVHKLTG.KVTSKNGQPIIPVLLPKYI.L 295

MBG  NNRQVVLAYANPTLSAV--------------                               303
EBO  DPVAPGDLTMVI.QDCDTCHSPASLPAVIEK                               326
(B)
```

Fig. 3. Comparison of the amino acid sequences of (A) the MBG and EBO VP35 proteins (for EBO VP35 shown is the sequence deduced starting from the first methionine of the open reading frame), and (B) the MBG and EBO VP40 proteins. Dots indicate that the amino acid is the same as that in MBG. Gaps, included to maximize alignment, are indicated by dashes.

←

Fig. 2. The fragment of the EBO genome (cDNA plus-strand) with the predicted amino acid sequences of the VP35 and VP40 proteins. The marks are identical to those in Fig. 1; the second methionine in the VP35 open reading frame is boxed.

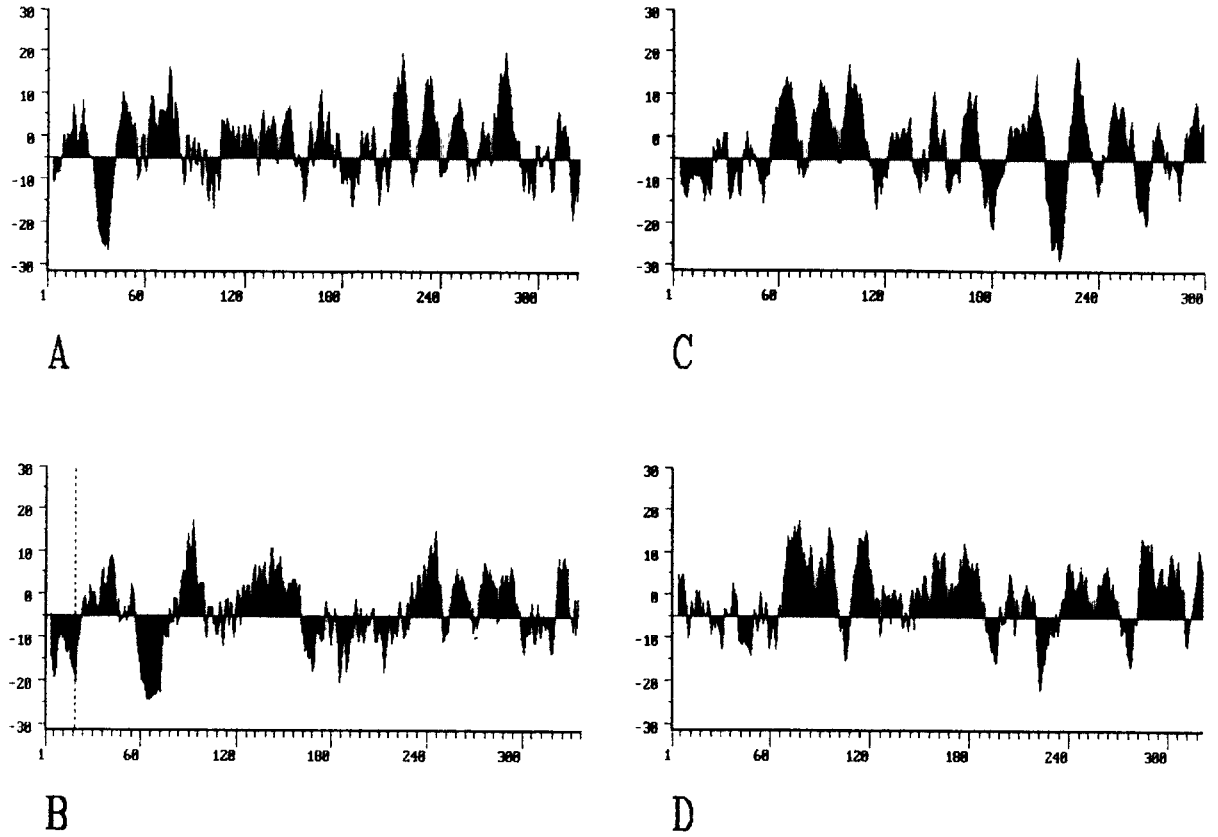


Fig. 4. The hydropathicity plots of the deduced polypeptide sequences, computed by the method of Kyte and Doolittle [13] using an interval of 9 aa. The hydrophobic regions are shaded above the midline, and hydrophilic ones below it. (A) The VP35 of MBG; (B) the VP35 of EBO; (C) the VP40 of MBG; (D) the VP40 of EBO. In B the position of the second AUG codon is shown by the dotted line.

REFERENCES

[1] Brown, F. (1989) Intervirology 30, 181–186.
[2] Martini, G. and Siegert, R. (1971) Marburg Virus Disease, Springer, New York.
[3] Kiley, M.P., Cox, N.J., Elliott, L.H., Sanchez, A., DeFries, R., Buchmeier, M.J., Richman, D.D. and McCormick, J.B. (1988) J. Gen. Virol. 69, 1957–1967.
[4] Sanchez, A. and Kiley, M.P. (1987) Virology 157, 414–420.
[5] Feldmann, H., Muhlberger, E., Randolf, A., Will, C., Kiley, M.P., Sanchez, A. and Klenk, H.-D. (1992) Virus Res. 24, 1–19.
[6] Sanchez, A., Kiley, M.P., Klenk, H.-D. and Feldmann, H. (1992) J. Gen. Virol. 73, 347–357.
[7] Muhlberger, E., Sanchez, A., Randolf, A., Will, C., Kiley, M.P., Klenk, H.-D. and Feldmann, H. (1992) Virology 187, 534–547.
[8] Sanchez, A., Kiley, M.P., Holloway, B.P., McCormick, J.B. and Auperin, D.D. (1989) Virology 170, 81–91.
[9] Volchkov, V.E., Blinov, V.M. and Netesov, S.V. (1992) FEBS Lett. 305, 181–184.
[10] Bukreyev, A.A., Kolichalov, A.A., Volchkov, V.E., Blinov, V.M., Netesov, S.V. and Sandakhchiev, L.S. (1991) Molekulyarnaya Genetika, Mikrobiologiya i Virusologiya 3, 24–30 (in Russian).
[11] Kozak, M. (1986) Cell 44, 283–292.
[12] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[13] Elliott, L.H., Kiley, M.P. and McCormick, J.B. (1985) Virology 147, 169–176.
[14] Bukreyev, A.A., Shestopalov, A.M., Kolichalov, A.A., Brovkin, A.I., Blinov, V.M. and Netesov, S.V. (1991) International Conference on Medical Biotechnology, Immunization and AIDS, Abstracts, R2-44, Leningrad, USSR, 1991.
[15] Volchkov, V.E., Chepurnov, A.A. and Netesov, S.V. (1991) International Conference on Medical Biotechnology, Immunization and AIDS, Abstracts, R2-45, Leningrad, USSR, 1991.