

# 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<sub>A</sub><sup>U</sup>CC<sub>A</sub><sup>U</sup>U<sub>A</sub><sup>G</sup>-UAAUU-5' and 3'-U<sub>A</sub><sup>A</sup>UUCUUUUU-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-

TCGAAGAATATTAAGGTTTCTTTAATATTCAGAAAAGGTTTTTATTCTCTCTTTCT 60  
TTTTGCAAACATATTGAAATAATAATTTTCAACAATGTGGGACTCATATATGCAACAA 120  
M W D S S Y M Q Q 9  
GTCAGTGAGGGTGTAGTACTGGAAAAGTCCCATAGATCAAGTGTGGTGGCCAAATCCC 180  
V S E G L M T G K V P I D Q V F G A N P 29  
TCAGAGAAGTTACACAAGAGAAGGAAACCAAAGGCACAGTTGGACTACAATGCAGCCCT 240  
S E K L H K R R K P K G T V G L Q C S P 49  
TGCTTAATGTCAAAGGCGACAAGCACTGATGATATTGTTTGGGACCAACTGATCGTGAAG 300  
C L H S K A T S T D D I V W D Q L I V K 69  
AAAACACTAGCTGACTTACTTATACCGATAAATAGGCAGATATCGGACATTCAAAGCACT 360  
K T L A D L L I P I N R Q I S D I Q S T 89  
CTAAACGAAGTAACAACAAGAGTCCATGAAATGAGCGGCAATTACATGAGATAACCCCA 420  
L N E V T T R V H E I E R Q L H E I T P 109  
GTGTTAAAAATGGGAAGGACACTGGAAGCAATTTCCAAGGGGATGTCAGAAAATGTTAGCC 480  
V L K M G R T L E A I S K G H S E M L A 129  
AAATACGACCACCTCGTAATTTCAACTGGAAGAACCCTGCACCAGCTGCTGCCTTTGAT 540  
K Y D H L V I S T G R T T A P A A A F D 149  
GCTTACTTAAATGAGCATGGTGTCCCTCCCCCAACCTGCGATTTTCAAAGATCTTGGG 600  
A Y L N E H G V P P P Q P A I F K D L G 169  
GTTGCTCAACAAGCTTGTAGTAAGGGGACCATGGTTAAAAATGAAACAACAGATGCAGCC 660  
V A Q Q A C S K G T H V K N E T T D A A 189  
GACAAGATGTCGAAAGTTCTTGAACCTAGTGAGGAGACGTTCTCCAAGCCAAATCTTCA 720  
D K M S K V L E L S E E T F S K P N L S 209  
GCTAAGGATTTAGCCCTTTTGTGTTTACCATCTACCCGGCAACAACACTCCATTCCAT 780  
A K D L A L L L F T H L P G N N T P F H 229  
ATCCTAGCTCAAGTCTTTCAAAAATGCTTACAAGTCAGGAAAGTCCGGAGCATTTTGG 840  
I L A Q V L S K I A V K S G K S G A F L 249  
GATGCATTTACCAGATTTAAGTGAAGGAGAGAATGCTCAGGCAGCATTGACTCGACTA 900  
D A F H Q I L S E G E N A Q A A L T R L 269  
AGCAGAACATTTGATGCTTCCCTCGGAGTAGTCTCCAGTGATAAGAGTCAAAAACCTC 960  
S R T F D A F L G V V P P V I R V K N F 289  
CAAAACAGTCCCTCGCCCATGTCAA AAAAGTCTTCGGGCTGTTCTCCCAACCAACAAT 1020  
Q T V P R P C Q K S L R A V P P N P T I 309  
GACAAAGGATGGGCTGTGTTTATTCATCTGAGCAAGGTGAGACCGGCCCTGAAAATC 1080  
D K G W V C V V S S E Q G E T R A L K I 329  
TAATTCCTATTGTTAACAGTTGCAGGGGGAGTGATCTTCCGAGTTGATACAAAAGCACT 1140  
\*  
AAACATTTCAAAGCATATATGTGGGCAAAACGTGACTAGACCATCTTAATAGAAGTAGT 1200  
AATTTATTCTGTCTTAAGTGTGATTTTCACTTGAAGAGTTAAATGGTGATAGATTA 1260  
TCCTTGAAGTAACTTTTATATATTATAGAGGAACATAATTAACAACAAGGGGTCT 1320  
ACCTAACAGGTATGACTGAGTATAGTATATTTTATAAACCAAGCAATTTGACTTCTCAC 1380  
TTTTTAAGAAATCAACTAACAACATAGAAAACATATTTATCCTTGTGTAATTTCTCGGCTA 1440  
GTTGGAATTAAGTCTTGTGCAATCAAGACGCTTATTCATAGTAGATTATATGATTTT 1500  
TATAAGTTTAAAGATATCTTAAATTATACCCACAAGAGATACTGTTTAAATTAAGAAAAAC 1560  
TATGAAGAACATTAAGAAGATCTTCTCCTAGTGTCTTTTACTGGAAGGAGTATCCC 1620  
AATCTCAGCTTGTGAATTAATTTACTTAAGTATCTTTTAAAAATTAATTCACACA 1680  
AGGTAGTTTGGGTTTATATCTAGAACAAATTTTAAATATGGCCAGTTCAGCAATTACAAC 1740  
M A S S S N Y N 8  
ACATACATGCAATACTTGAACCCCTCCTTATGCTGATCACGGTGCAAACCAAGTGGATC 1800  
T V H Q V L N P P P V A D H G A N Q L I 28  
CGGGCGGATCAGCTATCAAATCAGCAGGGTATAACTCCAATTTATGTGGGTGACTTAAAC 1860  
P A D Q L S N Q Q G I T P N V V G D L N 48  
CTAGATGATCAGTTCAAAGGGAATGCTGCCATGCTTCACTTTAGAGGCAATAATTGAC 1920  
L D D Q F K G N V C H A F T L E A I I D 68  
ATATCTGCGTATAATGAACCAACAGTCAAAGGTGTTCCAGCATGGCTGCCTCTCGGGATT 1980  
I S A V N E P T V K G V P A W L P L G I 88  
ATGAGCAATTTGAATATCCTTTAGCTCATACTGTGGCTGCGTTGCTCACAGGCAGCTAT 2040  
H S N F E V P L A H T V A A L L T G S Y 108  
ACAATCACCAATTTACTCATAATGGGCAAAAATTCGTCCGTGTAATTCGACTCGGTACA 2100  
T I T Q F T H N G Q K F V R V N R L G T 128  
GGAATCCCAGCACCCACTCAGAATGTGCGTGAAGGAAATCAAGCTTTTATTGAGAAT 2160  
G I P A H P L R M L R E G N Q A F I Q N 148  
ATGGTATCCCCAGAAAATTTTCCACTAATCAATTCACCTACAATCTCACTAACTTAGTA 2220  
M V I P R N F S T N Q F T V N L T N L V 168  
TTGAGTGTGAAAAGCTTCCCTGATGATGCTGGCGCCATCCAAGGACAAATTAATTGGG 2280  
L S V Q K L P D D A W R P S K D K L I G 188  
AACACCATGATCCCGCAGTCTCCATACACCCGAATTTGCCACCCATTGTTCTACCAACA 2340  
N T M H P A V S I H P N L P P I V L P T 208  
GTCAAGAAGCAGGCTTATCGTCAGCATAAAAATCCCAACAATGGACCCTGCTGGCCATA 2400  
V K K Q A V R Q H K N P N N G P L L A I 228  
TCTGCAATCCTTCAACACTGAGGTCGAGAAGTCCCAGAGAAGCAGACAGCTGTTTAGG 2460  
S G I L H Q L R V E K V P E K T S L F R 248  
ATTTCACTTCCGCGATATGTTCTCAGTAAAGAAGGTATGATGAAGAAAAGGGGAGAA 2520  
I S L P A D M F S V K E G M M K K R G E 268  
AATTCCTCCGTTGTTTATTTCAAGCACCTGAGAACTTCCCTTTGAATGGCTTCAACAAC 2580  
N S P V V Y F Q A P E N F P L N G F N N 288  
AGACAAGTTGACTAGCGTATGCGAATCCAAGCTCAGTCCGCTTTGAAAATAAGTCTCAA 2640  
R Q V V L A V A N P T L S A V \* 303  
ATGAGACAGGAGTCCATCTGCATAAGAAGCATGGCCATAATGGGTGCTGTTAAGTTCTC 2700  
ACAAGATTAGTTGATTTGATTTCAATAATGCTTTAACCTTACATGCTGCTTTAAATGG 2760  
TTAATTAAGCTGATCAGCTTGAAGATGTAATCTCTTTGGGTATCAGATCTATAATGG 2820  
GTTTACTAGATTATATAAAGAAAATAGTAATGTTTATAAACAATCTTGCTTAGTTTTA 2880  
CTTTGATTTACTAACATATATCAATGTCGCCCTTTCATTGCTAAGTAAACTCAACTGATGAT 2940  
GATATTCCTTCTGAAATAGTAAGAAAAA 2968

ties. The latter corresponds to the transcriptional stop signal and mRNAs extremity sequence 3'-UAAUUC-UUUUU-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'-GCU<sup>Δ</sup>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

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

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.

←

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.

ATGATGAAGATTAAAACCTTCATCATCCTTACGTCAATTGAATTCTCTAGCACTCGAAGC 60  
TTATTGTCTTCAATGTAAGAAAAAGCTGGTCTAACAAGATGACAAC TAGAACAAAAGGGC 120  
M T T R T K G 7  
AGGGGCCATACTGCGGCCACGACTCAAACGACAGAATGCCAGGCCCTGAGCTTTCGGGC 180  
R G H T A A T T Q N D R [H] P G P E L S G 27  
TGGATCTCTGAGCAGCTAATGACCGGAAGAATTCCTGTAAAGCAGACATTTCTGTGATATT 240  
W I S E Q L M T G R I P V S D I F C D I 47  
GAGAACAATCCAGGATTATGCTACGCATCCCAAATGCAACAACGAAGCCAAACCCGAAG 300  
E N N P G L C V A [S Q M Q Q T K P N P K] 67  
ACGCGCAACAGTCAAACCCAAACGGACCCAATTTGCAATCATAGTTTTGAGGAGGTAGTA 360  
T R N S Q T Q T D] P I C N H S F E E V V 87  
CAAACATTGGCTTCATTGGCTACTGTTGTGCAACAACAACCATCGCATCAGAATCATT 420  
Q T L A S L A T V V Q Q Q T I A S E S L 107  
GAACAACGCATTACGAGTCTTGAGAATGGTCTAAAGCCAGTTTATGATATGGCAAAAACA 480  
E Q R I T S L E N G L K P V V D M A K T 127  
ATCTCCTCATTGAACAGGGTTGTGCTGAGATGGTTCGAAAATATGATCTTCTGGTGATG 540  
I S S L N R V C A E M V A K V D L L V M 147  
ACAACCGGTGCGGCAACAGCAACCGCTGCGGCAACTGAGGCTTATTGGGCGAACATGGT 600  
T T G R A T A T A A A A T E A V W A E H G 167  
CAACCACCACCTGGACCATCACTTTTGAAGAAAGTGGGATTTCGGGGTAAGATTGAATCT 660  
Q P P P G P S L V E E S A I R G K I E S 187  
AGAGATGAGACCGTCCCTCAAAGTGTAGGGAGGCATCAACAATCTAAACAGTACCACT 720  
R D E T V P Q S V R E A F N N L N S T T 207  
TCACTAAGTGAAGAAAATTTGGGAAACCTGACATTCGGCAAAAGGATTGAGAAACATT 780  
S L T E E N F G K P D I S A K D L R N I 227  
ATGTATGATCACTTGCTGGTTTGGAACTGGCTTCCACCAATAGTACAAGTATTTGT 840  
M V D H L P G F G T A F H Q L V Q V I C 247  
AAATTGGGAAAAGTAGCAACTCATTGGACATCATGCTGAGTTCAGGCCAGCCTG 900  
K L G K D S N S L D I I H A E F Q A S L 267  
GCTGAAGGAGACTCTCTCAATGTGCCCTAATTCAAATTACAAAAGAGTTCCAATCTTC 960  
A E G D S P Q C A L I Q I T K R V P I F 287  
CAAGATGCTGCTCCACCTGTGATCCACATCCGCTCTCGAGGTGACATTCGCCGAGCTTG 1020  
Q D A A P P V I H I R S R G D I P R A C 307  
CAGAAAAGCTTGCGTCCAGTCCCACCATCGCCCAAGATTGATCGAGGTGGGTATGTGTT 1080  
Q K S L R P V P P S P K I D R G W V C V 327  
TTTCAGCTTCAAGATGGTAAAACACTTGGACTCAAAAATTTGAGCCAATCTCCCTCCCTC 1140  
F Q L Q D G K T L G L K I \* 340  
CGAAAAGGCGAATAATAGCAGAGGCTTCAACGGCTGAACATAAGGGTACGTTACATTA 1200  
TGATACACTTGTGAGTATCAGCCCTGGATAATATAAGTCAATTAACGACCAAGATAAAA 1260  
TTGTTTCAATCTCGCTAGCAGCTTAAAAATATAATGTAATAGGAGCTATATCTCTGACAG 1320  
TATTATAATCAATTGTTAATAAGTAACCCAAACAAAAGTGAAGATTAAAGAAAAACC 1380  
TACCCTCGGCTGAGAGAGTGTTTTTTCATTAACCTTCATCTTGTAAACGTTGAGCAAAAT 1440  
GTTAAAAATATGAGGCGGGTTATATTGCCTACTGCTCCTCTGAAATATATGAGGCCATA 1500  
M R R V I L P T A P P E Y M E A I 17  
TACCCTGTGAGGTCAAATTCACAAATGCTAGAGGTGGCAACAGCAATACAGGCTTCTCTG 1560  
V P V R S N S T I A R G G N S N T G F L 37  
ACACCGGAGTCAAGTCAATGGGGACACTCCATCGAATCCACTCAGGCCAATTGCCGATGAC 1620  
T P E S V N G D T P S N P L R P I A D D 57  
ACCATCGACCATGCCAGCCACACACAGGAGTGTGTCATCAGCATTCCATCCTTGAAGCT 1680  
T I D H A S H T P G S V S S A F I L E A 77

ATGGTGAATGTATATCGGGCCCCAAAGTGTACTAATGAAGCAAATTCAAATTTGGCTTCT 1740  
M V N V I S G P K V L M K Q I P I W L P 97  
CTAGGTGTGCTGATCAAAGACCTACAGCTTTGACTCAACTACGGCCGCCATCATGCTT 1800  
L G V A D Q K T Y S F D S T T A A I M L 117  
GCTTACATACATATACCCATTTCGGCAAGGCAACCAATCCACTTGTGAGAGTCAATCGG 1860  
A S V T I T H F G K A T N P L V R V N R 137  
CTGGGTCTGGAATCCCGGATCATCCCTCAGGCTCCTGCGAATTGGAACACAGGCTTTC 1920  
L G P G I P D H P L R L L R I G N Q A F 157  
CTCCAGGAGTTCGTTCTCCGCCAGTCCAACCTACCCAGTATTTACCTTTGATTTGACA 1980  
L Q E F V L P P V Q L P Q Y F T F D L T 177  
GCATCAAACCTGATCAACCAACCTGCCTGCTGCAACATGGACCGATGACACTCCAACA 2040  
A L K L I T Q P L P A A T W T D D T P T 197  
GGATCAAATGGAGCGTTGCGTCCAGGAATTTCAATTCATCCAAAACCTGCGCCCATTTCT 2100  
G S N G A L R P G I S F H P K L R P I L 217  
TTACCCAACAAGTGGGAAGAAGGGGAACAGTCCGATCTAACATCTCCGGAGAAAAATC 2160  
L P N K S G K K G N S A D L T S P E K I 237  
CAAGCAATAATGACTTCACTCCAGGACTTTAAGATCGTTCCAATTTGATCCAACAAAAAT 2220  
Q A I M T S L Q D F K I V P I D P T K N 257  
ATCATGGGAATCGAAGTCCAGAACTCTGGTCCACAAGCTGACCGGTAAGAAGGTGACT 2280  
I M G I E V P E T L V H K L T G K K V T 277  
TCTAAAAATGGACAACCAATCATCCTGTTCTTTTGCCAAAGTACATTTGGGTTGGACCCG 2340  
S K N G Q P I I P V L L P K V I G L D P 297  
GTGGCTCCAGGAGACTCAACATGGTAATCACACAGGATTGTGACACGTGATCTTCCCT 2400  
V A P G D L T M V I T Q D C D T C H S P 317  
GCAAGTCTTCCAGCTGTGATTGAGAAGTAATTGCAATAATTGACTCAGATCCAGTTTTAT 2460  
A S L P A V I E K \* 326  
AGAATCTTCTCAGGGATAGTGATAACATCTATTAGTAATCCGTCCATAGAGGAGACAC 2520  
TTTTAATGATCAATATACTAAAGGTGCTTACACCATTTGCTTTTTTCTCTCTAAATG 2580  
TAGAACTTAACAAAAGACTCATAATATACTTGTTTTAAAGGATTGATTGATGAAAGATC 2640  
ATAACTAATAACATTACAAATAATCCTACTATAATCAATACGGTGATTCAAATGTTAATC 2700  
TTTCTCATGACATACTTTTTGCCCTTATCCTCAAATTCGCTGATGCTTACATCTGAG 2760  
GATAGCCAGTGTGACTTGGATTGGAATGTGGAGAAAAAATCGGGACCCATTTCTAGGTT 2820  
GTTCAAAATCCAAGTACAGACATTGCCCTTCTAAATTAAGAAAAA 2865

MBG	-----HWSSVMQQVSEGLMTGKVPIDQVFGANPSEKLHRRKPKG	41
EBO	MTTRTKGRGHTAATTQNDR.PPELSGWI..Q..RI.VSDI.CDIENNPGLCYASQM	60
MBG	TVGLQCSPLMSKATSTDDIVWDGLIVKKTADLLIPINRQISDIQSTLNEVTTTRVHEIE	101
EBO	QTKPNPKTRNSQTQ.D-PICNHSFEE.VQ..S.ATVVQG.TIASE.LEGRI.S-----	113
MBG	RQLHEITPVLKMGRTLEAISKGMSEMLAKVDHLVISTGRRTAPAAAFDAYLNEHGVPVPPG	161
EBO	-LEGLK..VD.AK.ISSLNRVCA..V...L..MT..A..T..TE..WA..G...G	172
MBG	PAIFKDLGVAQQACSKGTMVKNETDDAADKMSKVLELSEETFSPNLSAKDLALLLFTHL	221
EBO	.SLYEESAIRGKIE.RDET.PGSVRE.FNNLNSTTS.T..N.G..DI.....RNIMVD..	232
MBG	PGNNTPFHILAQVLSKIAVKSCKSGAFLDAFHQI-LSEGENAQAALTRLSRTFDFLGVV	280
EBO	..FG.A..G.V..IC.LG-DSN.LDIH.EF.AS.A..DSP.C..IQITKRVPI.GDAA	291
MBG	PPVIRVKNFQTVPRPCQKSLRAVPPNPTIDKGWVCVVSSEOGETRALKI	329
EBO	....HIRSRGDI..A...P...S..K..R....FQLGD.K.LG...	340
(A)		
MBG	MASS---SNYNTYMQYLNPPPYADHGA---NGLIPADQLSNGGGITPNYVGDNLNDD---	51
EBO	RRVILPTAPPE..EAIY.VRSNSTI.RGG.SNTGFLTPESVN.D..SNPLRPIA..TID	60
MBG	---QFKGNVCHAFTLEAIIIDISAVNEPTVKGVPAWLPLGIMSNFEYPLAHTVAALLTGSY	108
EBO	HASHTP.S.SS..I..MVNVISGPKVLM.QI.I...VADQKT.SFDS.T..IMLA..	120
MBG	TITQFTHNGQKFVRVNRRLGTGIPAHPLRMLREGNAFIONMVI PRNFSTNOFTVNLTLV	168
EBO	...H.GKATNPL.....P...D...L..I.....L.EF.L.PVQLPGV..FD..A.K	180
MBG	LSVGKLPDDAWRPSKDKLIGNTHHPAVSIHPNLPPIVLP--TVKKQAVRQHKNPNGPLL	226
EBO	.IT.P..AAT.TDDTPTGSGALR.GI.F..K.R..L..NKSG..GNSADLTS.EK----	236
MBG	AISGILHQLRVEKVPEKTSLFRISLPADMFSVKEGMMKKRGENSPVVYFQAPENFPLNGF	286
EBO	..GA.MTS.GDF.IVPIDPTKN.MGIEVPETLVHKLGTG.KVTSKNGGPIIPVLLPKVI.L	295
MBG	NNRQVVLAYANPTLSAV-----	303
EBO	DPVAPGDLTHVI.GDCDTCHSPASLPVIEK	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.

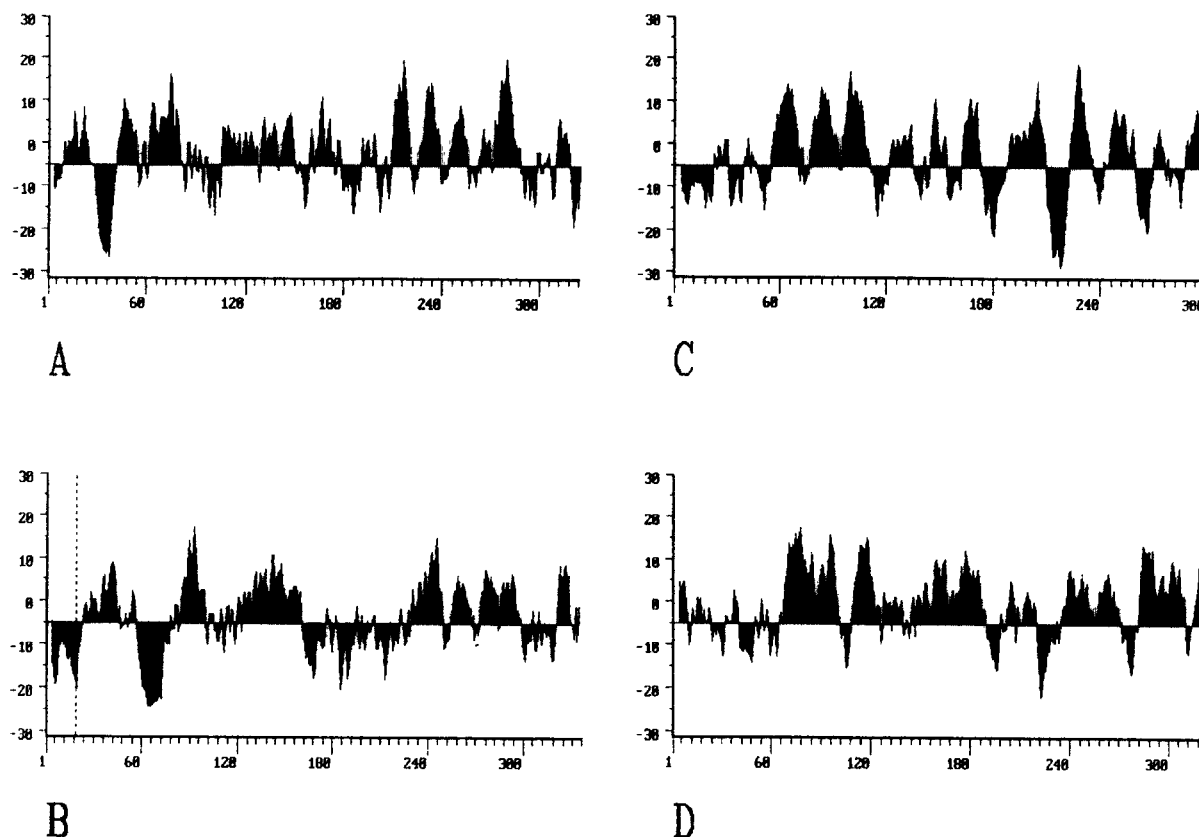


Fig. 4. The hydropathicity plots of the deduced polypeptide sequences, computed by the method of Kyte and Doolittle [13] using an interval of 9 aa. The hydrophobic regions are shaded above the midline, and hydrophilic ones below it. (A) The VP35 of MBG; (B) the VP35 of EBO; (C) the VP40 of MBG; (D) the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