

Marburg Virus Gene 4 Encodes the Virion Membrane Protein, a Type I Transmembrane Glycoprotein

CHRISTIANE WILL,¹ ELKE MÜHLBERGER,¹ DIETMAR LINDER,² WERNER SLENCZKA,¹
HANS-DIETER KLENK,¹ AND HEINZ FELDMANN^{1†*}

*Institut für Virologie, Philipps-Universität, 3550 Marburg,¹ and Biochemisches Institut,
Justus-Liebig-Universität, 6300 Giessen,² Germany*

Received 4 September 1992/Accepted 3 December 1992

Gene 4 of Marburg virus, strain Musoke, was subjected to nucleotide sequence analysis. It is 2,844 nucleotides long and extends from genome position 5821 to position 8665 (EMBL Data Library, emnew: MVREPCYC [accession no. Z12132]). The gene is flanked by transcriptional signal sequences (start signal, 3'-UACUUCUUGAAAU-5'; termination signal, 3'-UAAUUCUUUUU-5') which are conserved in all Marburg virus genes. The major open reading frame encodes a polypeptide of 681 amino acids (M_r , 74,797). After *in vitro* transcription and translation, as well as expression in *Escherichia coli*, this protein was identified by its immunoreactivity with specific antisera as the unglycosylated form of the viral membrane glycoprotein (GP). The GP is characterized by the following four different domains: (i) a hydrophobic signal peptide at the amino terminus (1 to 18), (ii) a predominantly hydrophilic external domain (19 to 643), (iii) a hydrophobic transmembrane anchor (644 to 673), and (iv) a small hydrophilic cytoplasmic tail at the carboxy terminus (674 to 681). Amino acid analysis indicated that the signal peptide is removed from the mature GP. The GP therefore has the structural features of a type I transmembrane glycoprotein. The external domain of the protein has 19 N-glycosylation sites and several clusters of hydroxyamino acids and proline residues that are likely to be the attachment sites for about 30 O-glycosidic carbohydrate chains. The region extending from positions 585 to 610 shows significant homology to a domain observed in the envelope proteins of several retroviruses and Ebola virus that has been suspected to be responsible for immunosuppressive properties of these viruses. A second open reading frame of gene 4 has the coding capacity for an unidentified polypeptide 112 amino acids long.

Marburg virus (MBG), the prototype of the filoviruses, causes a severe hemorrhagic disease in humans, with mortality rates of up to 35%. Filovirus infections were unknown until 1967, when MBG was first isolated in outbreaks among laboratory workers in Europe (Germany and Yugoslavia). The source of these simultaneous outbreaks was infected African green monkeys (*Cercopithecus aethiops*) imported from Uganda (20). Since then, sporadic, virologically confirmed MBG disease cases have occurred in various parts of Africa (9, 17, 34). The natural reservoir for human and other primate infections with MBG, as well as any other filovirus, is still unknown. Characterization of several MBG isolates showed that the geographically and temporally distinct outbreaks have been due to genetically distinct but antigenetically related virus strains (17).

MBG particles are composed of a helical nucleocapsid surrounded by a lipid envelope. The genome is 19.1 kb long and consists of a single molecule of nonsegmented negative-strand RNA (7, 16, 25). It shows a linear arrangement of genes (3' untranslated region-nucleoprotein [NP]-VP35-VP40-glycoprotein [GP]-VP30-VP24-RNA-dependent RNA polymerase [L protein]-5' untranslated region) which encode seven structural proteins: the L protein (267 kDa), the GP (170 kDa), the NP (96 kDa), and four additional proteins (VP40 [38 kDa], VP35 [32 kDa], VP30 [28 kDa], and VP24 [24 kDa]) (7). The nucleotide sequences of the NP (29) and the L protein (21) have been elucidated.

The GP has an M_r of 170,000 when analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) and is present in the viral membrane as a homotrimer (8). The oligosaccharide structures of the highly glycosylated protein include oligomannosidic and hybrid type N-glycans, as well as bi-, tri-, and tetraantennary complex species. In addition, large amounts of neutral mucin-type O-glycans (type 1 and 2 core structures) are present. Sialic acid residues are absent on both O- and N-glycans (8, 11). Since the GP is the only membrane protein exposed on the viral surface, it is reasonable to assume that it is responsible for receptor binding and membrane fusion and therefore plays a key role in entry of the virus into cells. It also is probably the main target for the immune response of an infected host.

Elucidation of the structure of the GP is therefore instrumental for the understanding of virus replication and pathogenesis. In the present study, we analyzed the primary structure of the GP by cloning, sequencing, and expression of the corresponding gene.

MATERIALS AND METHODS

Cell line and virus. The Musoke strain of MBG (34) was used. The virus was propagated in E6 cells, a cloned Vero cell line (ATCC CRL 1586).

Genomic RNA (vRNA) and subgenomic RNA (mRNA) isolation. For vRNA isolation, E6 cells were infected with MBG at a multiplicity of infection of 10^{-2} PFU per cell. Growth and purification of virus and vRNA isolation were performed as described previously (8, 21). Virus-specific mRNA was obtained from E6 cells infected at multiplicities

* Corresponding author.

† Present address: Special Pathogens Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control, 1600 Clifton Road, Mailstop G14, Atlanta, GA 30333.

of infection of 1 to 10 PFU per cell. Isolation and purification were performed as described elsewhere (7).

cDNA synthesis, cloning, screening, and sequencing. cDNA was synthesized from vRNA and mRNA by using random hexamers [oligo-pd(N6)], as well as specific oligonucleotides for priming (7, 12, 21). cDNA was supplied with *Bam*HI adaptors and ligated into the *Bam*HI sites of plasmids pUEX1 (3) and pUC18. Recombinant cDNA clones were screened for viral specificity by colony hybridization and confirmed by Southern blot hybridization using ³²P-labeled vRNA. DNA sequence analysis was performed by a protocol based on the chain termination method (30). Several parts of the sequence were difficult to interpret because of band compression in G-C-rich areas. To overcome such problems, we used the guanine analog inosine (c⁷dITP) instead of dGTP (Deaza T7 Sequencing Mixes; Pharmacia). In addition, sequence data were confirmed by direct sequencing of vRNA and, to a lesser extent, mRNA (32).

Expression of gene 4 coding region. The entire open reading frame (ORF) of the MBG GP-encoding gene was synthesized from purified vRNA by the polymerase chain reaction (PCR) technique (7, 21). The following oligonucleotides, containing *Sal*I restriction endonuclease sites (underlined) at their 5' ends, were used for amplification: 5'-CAGGTCGACACCTAACATGAAGACCACATGTTTCC-3' (mRNA sense, bases 5931 to 5957) and 5'-CAGGTCGACGGGCTCAGTAGCAATTGTCCTAAAGTC-3' (vRNA sense, bases 8032 to 8006). The PCR-amplified product was digested with *Sal*I and ligated into the *Sal*I sites of plasmids pGEM3Zf(+) (Promega) and pQE (Diagen). In vitro transcription [plasmid pGEM3Zf(+)] and translation were performed as described previously (7). The [³⁵S]methionine-labeled translation products were immunoprecipitated by using a GP-specific rabbit antiserum raised against the purified GP (see below), subjected to SDS-PAGE, and detected by fluorography. Unlabeled products were detected by immunoblotting. For expression in a prokaryotic system, plasmid pQE (Diagen) and *Escherichia coli* M15 were used. M15 carries plasmid pREP4, which contains the gene for neomycin phosphotransferase (kanamycin resistance) and the *lac*I gene that encodes the *lac* repressor. Expression of the gene 4 ORF was induced by isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration of 2 mM for 2 h at 37°C. Cells were pelleted, washed, and lysed in 2× SDS-sample buffer. Proteins were subjected to SDS-PAGE analysis and blotted onto nitrocellulose membranes. Expression product of gene 4 was detected by using both anti-MBG guinea pig serum and the above-described GP-specific rabbit antiserum.

Purification of GP from virion particles. Purified MBG proteins were separated on preparative SDS-8% PAGE and stained with Coomassie brilliant blue. The GP was cut out, and the gel slices were placed into a BIOTRAP (Schleicher & Schuell) which was set into a horizontal electrophoresis chamber. Electroelution was performed in buffer containing 100 mM glycine, 20 mM Tris base, and 0.01% SDS for 16 to 20 h at 4°C and a constant voltage (200 V). The eluate was removed and concentrated by using Centricon-100 microconcentrators (Amicon). An aliquot of the concentrated eluate was electrophoresed by SDS-10% PAGE and analyzed by Coomassie brilliant blue staining, immunoblotting, and detection by the lectin peanut agglutinin (8). The purified GP was used directly for amino acid sequence analysis and immunization of rabbits.

Amino acid sequence analysis. For N-terminal sequence analysis, the combined GP was further purified by reversed-phase high-performance liquid chromatography on a column

(0.46 by 25 cm) of WP 300, C₄, 5 μm (Shandon, Astmoor, United Kingdom) using an acetonitrile gradient (0% B to 100% B in 40 min; A, 0.1% trifluoroacetic acid-10% acetonitrile; B, 0.1% trifluoroacetic acid-90% acetonitrile) at 60°C with a flow rate of 1 ml/min. The protein fractions collected at 63% acetonitrile were combined to obtain sufficient material (50 to 100 pmol) for amino acid sequence analysis (13). The GP was placed on a Biobrene plus (Applied Biosystems, Foster City, Calif.)-pretreated glass fiber disc and mounted in the cartridge of a 477A Protein/Peptide Sequencer (Applied Biosystems). Sequence determination was performed by using the standard protocol given by the manufacturer (normal-1). Identification of phenylthiohydantoin amino acid derivatives was performed by a 120A on-line phenylthiohydantoin analyzer (Applied Biosystems).

Computer analyses. Computer analyses were performed by using the HUSAR (Heidelberg Unix Sequence Analysis Resources) program, release 2.0 (German Cancer Research Center Heidelberg and Center for Molecular Biology, University of Heidelberg, 1990). For homology searching, the following sequencing data bases were used: GenBank release 72.0, SWISS.PROT release 22.0, PIR.PROT release 32.0, EMBL.NUCL release 30.0, and EMBL.Viral release 30.0.

GenBank-EMBL accession codes. The nucleotide sequence of the MBG genome is in the EMBL data library, emnew: MVREPCYC (accession no. Z12132). The GenBank accession codes for the amino acid sequences of the envelope proteins of the following retroviruses: SRV-1, VCLJSA; MPMV, VCLJMP; RSV, VCFVER; BAEV, JT0262; ASV, VCFVUR; ARV, VCV DAR; HTLV-I, VCLJCN; GaLV, VCLJGL; MoLV, VCVWEM; MCFFV, VCVWM1; KMLV, A03985; HTLV-II, VCLJT2; FeLV, VCVWGF; FeSV, VCMVSS. For definitions of abbreviations, see the legend to Fig. 5.

RESULTS

Gene 4 has a large ORF that encodes a polypeptide 681 amino acids long. Recombinant clone 37 hybridized to vRNA, demonstrating its viral specificity, and to three different mRNA transcripts, one which migrated with the size of 28S rRNA and two with approximately the size of 18S rRNA (data not shown). This clone did not hybridize to MBG NP-specific clones (29) or to L-protein-specific clones (21) and thus was selected for sequence analysis. The cDNA insertion of clone 37 is 4,179 nucleotides long and covers the genome from nucleotide positions 5430 to 9609. The clone contains all of gene 4 (the GP) and parts of the upstream (gene 3; VP40) and downstream (gene 5; VP30) genes (Fig. 1A and 2).

The entire nucleotide sequence of gene 4 is presented in Fig. 2 as mRNA positive-sense DNA. Including the final five A residues which belong to the GP mRNA poly(A) tail, the GP gene is 2,844 nucleotides long and extends from positions 5821 to 8665 of the viral genome (Fig. 1A and 2). The gene is flanked by signals for transcription, with the start signal 3'-UACUUCUUGUAAUU-5' (genomic negative sense) located at its 3' end (positions 5821 to 5834) and the termination (polyadenylation) signal 3'-UAAUUCUUUUU-5' (genomic negative sense) at its 5' end (positions 8655 to 8665).

The first AUG codon is located within the transcription start signal at the very 5' end of the mRNA transcript of gene 4 (genome nucleotide positions 5821 to 5823) and opens a short ORF of 117 nucleotides (Fig. 1B and 2). Both findings

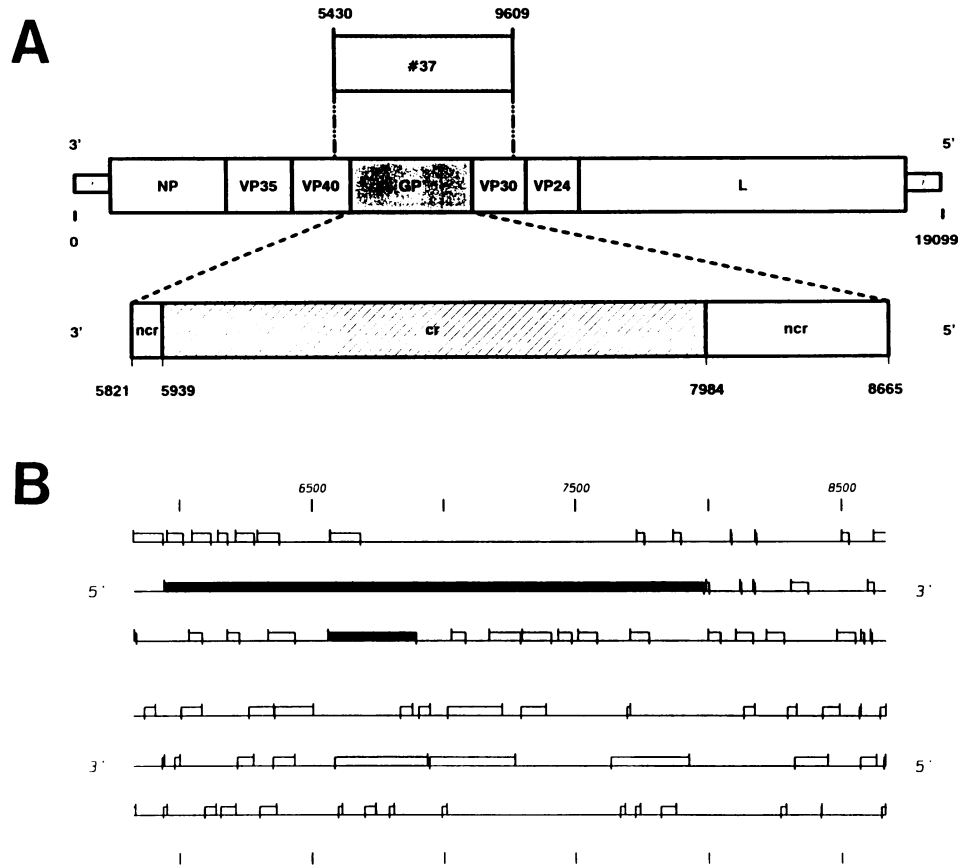


FIG. 1. Scheme of the gene that encodes the GP. (A) Genome localization. The figure shows that gene 4 of the linear genome encodes the GP. It further shows the lengths of the 3' and 5' noncoding regions (ncr) and the coding region (cr, ORF) of the GP-encoding gene. Clone 37 covers the entire GP-encoding gene and partial sequences of the upstream (VP40) and downstream (VP30) adjacent genes. (B) Coding capacity. The ORFs of viral complementary (m)RNA (top three frames) and vRNA (bottom three frames) are shown. All possible ORFs starting with the first AUG codon and terminated by a stop codon (AUG codons in between are not counted) are indicated by bars. The major ORF (5939 to 7984) and the second ORF (6558 to 6896) are marked by black bars.

suggest that this AUG codon is probably not used for initiation of protein synthesis. The next AUG codon (genome nucleotide positions 5939 to 5941) opens an ORF of 2,045 nucleotides which is terminated by the codon UAA at nucleotide positions 7982 to 7984 (Fig. 2). This AUG start codon is in a favorable context to serve as an initiation codon for eukaryotic ribosomes (A-AUG) (19), and the following ORF would encode a polypeptide of 681 amino acids. Consequently, the first 118 nucleotides at the 3' end (genome nucleotide positions 5821 to 5938) and the last 681 at the 5' end (genome nucleotide positions 7985 to 8665) (Fig. 1A and 2) of the gene 4 mRNA transcript are untranslated.

Gene 4 has the coding potential for a second polypeptide. Figure 1B demonstrates the coding potential of gene 4. All possible ORFs starting with the first AUG codon and terminated by a stop codon (AUG codons in between are not counted) in the viral complementary sense (positive-sense mRNA), as well as genomic negative-sense RNA, are indicated by bars. The largest ORF is present in the viral complementary sense and corresponds to the above-described ORF, starting at positions 5939 to 5941 and terminating at positions 7982 to 7984 (Fig. 1B and 2). An additional, smaller ORF is located in the viral complementary sense (AUG, nucleotide positions 6558 to 6560; UGA, 6894 to 6896) (Fig. 1B and 2) intersecting with the large ORF. The

AUG codon shows a good context for initiation of eukaryotic protein synthesis (19), and the started ORF would encode a polypeptide of 112 amino acids with an M_r of about 12,500. Two ORFs which would encode polypeptides with M_r s above 10,000 are located in the genomic negative-sense RNA (Fig. 1B). However, additional transcription signals which flank all MBG genes at their 3' and 5' ends are not present in negative- or positive-sense gene 4 RNA. Moreover, sequences favorable for RNA editing (4) or translational frameshifting are not found (15).

Expression of the gene 4 coding region. The coding region of gene 4, starting with the AUG codon at positions 5939 to 5941 and terminating at positions 7982 to 7984 (Fig. 1 and 2), was transcribed from vRNA by PCR into cDNA which was ligated into vector pGEM3Zf(+). 32 P-labeled recombinant plasmid DNA [pGEM3Zf(+)-gene 4 ORF] was hybridized to Northern (RNA) blots of purified MBG vRNA and mRNA of infected E6 cells. The PCR product identified the vRNA and a single virus-specific mRNA transcript that comigrated with the 28S rRNA (data not shown). Run-off transcripts were produced from linearized plasmid DNA by using T7 RNA polymerase. Translation products were labeled with [35 S]methionine during in vitro synthesis, immunoprecipitated with GP-specific rabbit antiserum, and subjected to SDS-PAGE (Fig. 3B). The largest and most prominent

```

ATGAGCAATTAATGCTGGTAAAGGTGAATTAATTCCTTAATTTGACAGAAATAATTTTGTCAAGTAAATATCTCATATCA 5910
TTGATTAAACAGAAAATACCTAAATGAAAGACAGATGTTTCTTATGCTTTATCTTAATTCAGAGCAAAAATATCCCAAT 6000
H K T T C G E L S L U L L Q G T K H L P F I 31
TTTAGAGTACTAGTAAATATCAACCCAAAATGGGATCGGTATGCTCCGGACCTCCAGAGAGAGAGAGAGTCCATCTGTGGG 6090
L E I A S N N Q P Q N V D S V C S G T L Q K T E D V H L M G 51
ATTCCAGCTGAGTGGCAAAAATGCTGATCCCTTTGGAGGATCCAGAGATGGGCTTTGGAGAGGTGTATCTCCAGAAATGT 6180
S T L F S G Q K V R D S P L E A S R K R A T F P F G V P P K 81
TGATACAGAGGGGGAGGCAAAATGCTACAAATATAGTGTACGGATCCCTGTGAAATGCTTTGTGTAGATGCTCCAC 6270
E Y T E G E E A R L C T C N N S V T D P P S G R S L L L D P P T 111
CAACCTCCCTGACTCCCAATGCAAAATCCATCCATCAATGAGTCAAAAGCCCTATCGCAGAGGGATCGCCCTCATATGAGGG 6360
N I R D Y P K C K T I H H I Q Q G N P H A Q G I A L H L W G 141
AGCATTTTTCGTATGATGATCCCTCCCAACAATGTACCGAGGCGAAGCTCTCACTGAGGAGCAATAGACAGCTATGATTGCA 6450
A F F P Y D R L D S T T M Y R G K V F P T E G M I A A M I V V N 171
TAAAGACTCCAGAAATATGTTTTCGCGAGAGCAAGGGTACCGGTATATGAATCTGACTTACTAATAAATTTGGACAGAG 6540
K T V H K H I F S R Q G G G G Y R H M N L T S T N K Y W T S S 201
TACCGAGCCAAAGCAATGACCTGATGTTTCGGCCCTTCAGAATACAACTCAAGAAGCAAAAGCAATGCTCCGTCAGAAAT 6630
N G T O T N D T C F G A L O E N Y R S T K N O T C A P S K K I 231
L H M T L D V S A A L F H N T T I L Q R T K H L L R P K T 251
ACTCCAGCAGCTCCGAGAGCCCTCCGAGATCACTCACAGACCCCAACTGATGCTCAACAATCTAATACCCAGCAGAGCA 6720
L H F C P F O V P R F I S H P S H P F H N T P T I S T N N T T D S S S U 261
TGATGATGAGCTCCGCAATCCGGCTCAGGCTCCGGAGACAGAACCCCAAGCAATCTGTGATGCTCCAGCAGAGAGGGCTTC 6810
D D E R T S A T S G S G S G E N E P P T T S L D M A V T I K O G L F S 291
M H N R T S Q H P A Q G P E N E N P T Q L L M R S P S K G F 311
ATCAGCAATCCGACCTCCCTCCAGCAGCAAGCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6900
Q C H P L P H H N O A R H S K E E T T O T I P K M L * 331
ACTAGCAAAATACAGCAATCCAGCAATCCCTCATACACTACCAATCTCTACTAGCAACCTCCCAAGCAACACT 6990
L D K H N T T R P S H P F H N T P T I S T N N T T S K R N T F 351
CAGCACTCTCTCCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAG 7080
S T L S A P F T N T N A A N T C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C 7080
A G C A C C T C C T C A G C A G C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C 7170
T T L P L P T G N P T T A K S T S S K K G P A T T A P N T T N 411
TGAGCAATCCAGCAGCTCCCTCCAGCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAAT 7260
E H F T S P E N E N D P S T A Q H L V Y F R R K N T T A G S L I L M R E 441
AGCAGCAATGCTCCCTTTTGGATGGTAAATAGCTCCCAATGATTTTGGAGCAGTCCCAATACAGAAAGCAATCTTGGAG 7350
G D M F P L D G D I N A P I D P V P N T K T I F D E S 471
CTCTAGTCTGCTCCCTCCAGCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAG 7440
S S S G A S A E E D O H A S P N I S L T L S Y F P N I N E N 501
CACTCCCTACCTCGAGAAATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7530
T A Y S D F E N E N D P S T A Q H L V Y F R R K N T T A G S L I L M R E 441
TTGATACCTGTTTTTGGCTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7620
W I P F P L G P G I E G L Y T A V L I K N Q N N L V C R L R R 561
TTAGCAATCCAGCAGCTCCCTCCAGCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAATCCAGCAAT 7710
L A N Q T A K S L E L L L R V T T E E R T F S L I N H R A I 591
TGATTTCTACACAGATGGGAGAGATCCAAAGTCTTGAGCCTGATTTGGCTCCGGATAGAGACTTGTCCAGCAATATTT 7800
P F L F E M P E F S E Y L E S F E S I G T E D L S R K N I 621
AVAGCAATTCAGCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7890
E Q I D Q I K K D E Q K E G T G W G L G G R W W T S D M G V 651
TCTTACTACCTGGGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7980
L T N L L L L L S I A V L I A L S C I C R I F T K Y I G 681
ATACGCTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8070
G A N C T G A T A C T A A N T G A T A C T A A N T G A T A C T A A N T G A T A C T A A N T G A T A C T A A N T 8160
T T C T A T T G A T G A T A C T A A N T G A T A C T A A N T G A T A C T A A N T G A T A C T A A N T G A T A C T 8250
T T A T T G A T G A T A C T A A N T G A T A C T A A N T G A T A C T A A N T G A T A C T A A N T G A T A C T A A N T 8340
A G A A A G C T G C T G T C A A A A G A G T C T T A A T A C C A C T A T T G C A A A A A T T A A T T A C T T T T C C G T A T A C T A G A T A A T 8430
C T T T G A C T G C A G C A T T A T A A C G A C T G C T C A A T T C A A T G C T G G A T G A A A T C T A A G A A A A A A G A G C C T C C C C C T A T T 8520
C T G G C T G A G A A A A T T C T T A T C C C C T A A A T G A C T G T A G G C T A G G C T C C T A A T C T G T T G A C G A T G A A T T G A T G 8610
A A T G A C A A A T G A T A G T G C A T T G T A T G A T C A A T T A C T T A T T A A G A A A A 8665

```

FIG. 2. Nucleotide sequence (mRNA positive-sense DNA) and deduced amino acid sequence of gene 4 of MBG, strain Musoke. The numbering of nucleotides is based on genome position (EMBL data library, emnew: MVREPCYC [accession no. Z12132]) (7). The transcription start and termination signals are underlined in the nucleotide sequence. Two strong hydrophobic regions, the signal peptide (1 to 18) and the transmembrane region (644 to 673), are underlined in the amino acid sequence. The deduced amino acid sequence of the second ORF is italicized in the single-letter code. A conserved domain (585 to 610) common to filoviral and retroviral envelope proteins is indicated by broken lines.

translation product that comigrated with the MBG NP on SDS-PAGE has an M_r of about 75,000, as expected for the totally deglycosylated GP. The same translation product was detected on immunoblots (data not shown).

The gene 4 coding region was further ligated into plasmid pQE for expression in *E. coli* M15. Expression was induced by addition of IPTG and detected on immunoblots by using the above-described GP-specific rabbit antiserum (data not shown) and anti-MBG guinea pig serum (Fig. 3A). Two protein bands were detected, of which the smaller and most prominent comigrated with the MBG NP and the above-described in vitro translation product of gene 4 (Fig. 3B). The second minor product, which migrated on SDS-PAGE with an M_r of about 150,000, probably represents an oligomeric form (dimer) of the gene 4 product expressed in this prokaryotic system. These data, together with the in vitro transcription and translation results, demonstrate that the large ORF of gene 4 encodes the MBG GP.

The gene product. The deduced amino acid sequence of the GP is shown in Fig. 2 beneath the nucleotide sequence and consists of 681 amino acids with a M_r of 74,797. It contains 10.3% acidic (Asp and Glu), 9.1% basic (Arg and

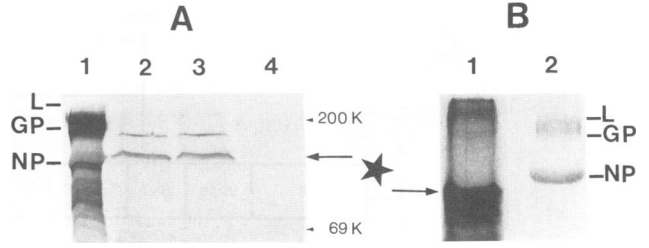


FIG. 3. Translation product of the gene 4 ORF. (A) Expression in *E. coli*. The coding region of gene 4 was synthesized from vRNA by PCR, ligated into plasmid pQE, and expressed in *E. coli*. Proteins were subjected to SDS-PAGE (10% polyacrylamide) and blotted onto nitrocellulose. Expression of the gene 4 ORF was monitored by using anti-MBG guinea pig serum. Lanes: 1, purified MBG proteins; 2, pQE-gene 4 coding region in *E. coli*; 3, pQE-gene 4 coding region in *E. coli*; 4, pQE in *E. coli*. (B) In vitro translation. The coding region of gene 4 was ligated into vector pGEM3Zf(+) and in vitro transcribed by using T7 RNA polymerase. The in vitro translation products were labeled with [³⁵S]methionine, immunoprecipitated by using GP-specific rabbit antiserum, subjected to SDS-10% PAGE, and detected by fluorography. Lanes: 1, [³⁵S]methionine-labeled in vitro translation products of the gene 4 ORF (pGEM3Zf); 2, [³⁵S]methionine-labeled purified virion proteins. *, deglycosylated form of the GP. The numbers to the right of lane 4 are molecular weights (10³).

Lys), 6.6% aromatic (Phe, Trp, and Tyr), and 25.9% hydrophobic (Ile, Leu, Met, Phe, Trp, Tyr, and Val) amino acid residues. The GP polypeptide has a net charge of -0.5 and an isoelectric point of 6.25. It shows a high content of asparagine, serine, and threonine residues (25%), resulting in the presence of 22 Asn sequons (Asn-X-Thr/Ser) (Asn positions 94, 171, 190, 202, 219, 223, 255, 310, 313, 325, 326, 337, 344, 345, 350, 360, 389, 408, 487, 564, and 619) for the attachment of N-glycosidically linked oligosaccharides and several clusters of three to four hydroxyamino acid residues (serine and/or threonine). These clusters and 19 (86%) of the Asn sequons are located in the middle of the polypeptide, between amino acids 170 to 500 (Fig. 2 and 4A). The polypeptide further contains 48 proline residues which contribute to a relatively high proline content (7.0%) for the protein. Most (35 residues) are located between amino acids 170 and 500 in several clusters of four to six, which makes the proline content for this part of the molecule unusually high (10.6%) (Fig. 2 and 4B). Thirteen cysteine residues are present in the amino acid sequence located at the N-terminal (1 to 170; four Cys residues) and C-terminal (500 to 681; seven Cys residues) ends of the polypeptide. However, they are nearly absent in the region between 170 and 500 (two Cys residues) (Fig. 2 and 4B). The deduced amino acid sequence shows two strong hydrophobic regions besides a few weaker ones. One is located at the N terminus of the polypeptide and extends from positions 1 to 18 (Fig. 2 and 4C) with eight hydrophobic, eight neutral, and two charged residues. The larger region is located close to the C terminus, at positions 644 to 673 (Fig. 2 and 4C) and contains 17 hydrophobic and 12 neutral residues and 1 charged residue. Hydrophilic regions are prominent within the middle of the molecule, from amino acid positions 180 to 520 (Fig. 2 and 4C). This region has also a relatively high antigenicity index (Fig. 4C).

To determine the precise amino terminus of the GP, MBG particles were isolated from the supernatant fluid of infected E6 cells 8 days postinfection. Virus was concentrated by centrifugation through a 20% sucrose cushion prior to purification by gradient centrifugation. This procedure mini-

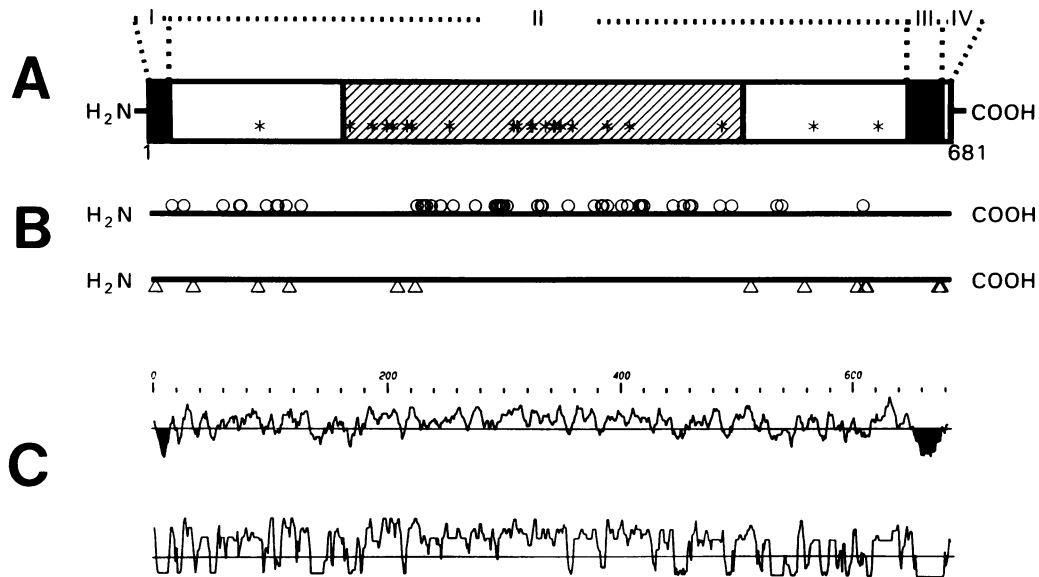


FIG. 4. Structure of the MBG GP. (A) Domain structure. The scheme shows the following domains of the protein: a hydrophobic signal peptide (1 to 18) (I), a predominantly external hydrophilic domain (19 to 643) (II), a hydrophobic transmembrane region (644 to 673) (III), and a short hydrophilic cytoplasmic tail (674 to 681) (IV). A hydroxyamino acid-rich domain with several clusters of serine and/or threonine residues is located within domain II (170 to 500, cross-striped). Potential N-glycosylation sites (Asn-X-Ser/Thr) are indicated by asterisks. (B) Locations of proline and cysteine residues. Proline residues are indicated by open circles, and cysteine residues are indicated by open triangles. (C) Hydrophobicity (upper part) and antigenicity (lower part) indices. Two strong hydrophobic regions are indicated by black areas. The smaller one at the N terminus represents the signal peptide (1 to 18), and the larger one represents the transmembrane region close to the C terminus (644 to 673). The region from about 180 to 520 shows a high antigenicity index (lower part).

mized the loss of membrane protein caused by the polyethylene glycol precipitation used in previous protocols (6, 21, 27). Viral proteins were separated on SDS-10% PAGE and stained with Coomassie brilliant blue. Under these conditions, the GP migrates as a diffuse band with an M_r of approximately 170,000 (8). The band was cut out, and the protein was purified by electroelution. The identity of the eluted GP was established by its reactivity with a monospecific antiserum and with the lectin peanut agglutinin (8) on nitrocellulose blots (data not shown). Following further purification by reversed-phase high-performance liquid chromatography, the protein was subjected to N-terminal amino acid sequence analysis. Leucine was identified as the unblocked amino-terminal amino acid residue, and the following 18 amino acid residues were determined in the order Pro-Ile-Leu-Glu-Ile-Ala-Ser-Asn-Asn-Gln-Pro-Gln-Asn-Val-Asp-Ser-Val-Cys (Fig. 2).

Comparison with the GP of EBO and envelope proteins of several retroviruses suggests the presence of an immunosuppressive domain at the C terminus of the MBG GP. The amino acid sequence of the MBG GP was compared with the C-terminal half (318 amino acid residues) of the glycoprotein of a related filovirus, Ebola virus (EBO), which was published briefly as a part of the glycoprotein sequence (36). Comparison revealed strong homology at the C termini (150 C-terminal amino acid residues) of both proteins, whereas the rest showed only limited homology (data not shown). Within the compared region, both polypeptides share several features: (i) seven highly conserved cysteine residues at the C terminus, (ii) a high content of asparagine, serine, and threonine residues, (iii) clusters of hydroxyamino acids, (iv) an unusually high proline content, (v) similar numbers of Asn sequons, (vi) a larger hydrophobic region of 30 amino acid residues at the C terminus, and (vii) a short hydrophilic

C-terminal tail (36). These data and the fact that both proteins are encoded by gene 4 of the linear arranged genomes clearly demonstrate their structural and functional similarity.

The EBO GP shares a conserved sequence of 26 amino acids with the envelope proteins of a number of retroviruses that is suspected to be responsible for immunosuppressive properties of these viruses (36), and this domain is also present in the MBG GP (Fig. 5; see Materials and Methods). It is located in the external part of the GP adjacent to the transmembrane anchor (positions 585 to 610) (Fig. 2 and 4A). There is 76.9% homology to EBO and 42.3 to 50.0% homology to the retroviruses in this domain (Fig. 5). All domains share a highly conserved N-terminal part (N-R-r-g-l-D) and three cysteine residues in identical positions at the less conserved C terminus (C-x₆-C-C).

DISCUSSION

In the present study, we elucidated the nucleic acid sequence of gene 4 of MBG. After *in vitro* translation and expression of the major ORF of this gene in *E. coli*, the product was clearly identified by its immunoreactivity with specific antisera as the viral GP. Analysis of the amino-terminal amino acid sequence of the GP also proved that this protein is encoded by gene 4. The GP-encoding gene is 2,844 nucleotides long and is located between the genes that encode the core proteins at the 3' end and the L-protein-encoding gene at the 5' end of the genome (7). This is in line with the localization of the EBO GP-encoding gene (36) and the envelope protein-encoding genes of other nonsegmented negative-strand RNA viruses which are located in the variable middle parts of the genomes (7, 24). Gene 4 is flanked by

MBG	---	L	I	N	R	H	A	I	D	F	L	L	T	R	W	-	G	G	T	C	K	V	L	G	P	D	C	C	---	100.0%	
EBO	---	I	L	N	R	K	A	I	D	F	L	L	Q	R	W	-	G	G	T	C	H	I	L	G	P	D	C	C	---	76.9%	
SRV-1	---	L	Q	N	R	R	G	L	D	-	L	L	T	A	E	Q	G	G	I	C	L	A	L	Q	E	K	K	C	---	50.0%	
MPMV	---	L	Q	N	R	R	G	L	D	-	L	L	T	A	E	Q	G	G	I	C	L	A	L	Q	E	K	K	C	---	50.0%	
RSV	---	L	Q	N	R	A	A	I	D	F	L	L	L	A	H	-	G	H	G	C	E	D	V	A	G	M	C	C	---	50.0%	
BAEV	---	L	Q	N	R	R	G	L	D	-	L	L	T	A	E	Q	G	G	I	C	L	A	L	Q	E	K	K	C	---	50.0%	
ASV	---	L	Q	N	R	A	A	I	D	F	L	L	L	A	H	-	G	H	G	C	E	D	I	A	G	M	C	C	---	50.0%	
ARV	---	L	Q	N	R	R	G	L	D	-	L	L	T	A	E	Q	G	G	I	C	L	A	L	Q	E	K	K	C	---	50.0%	
HTLV-I	---	A	Q	N	R	R	G	L	D	-	L	L	F	W	E	Q	G	G	L	C	K	A	L	Q	E	Q	C	C	---	46.2%	
GALV	---	L	Q	N	R	R	G	L	D	-	L	L	F	L	K	E	G	G	L	C	A	A	L	K	E	E	C	C	---	46.2%	
MoLV	---	L	Q	N	R	R	G	L	D	-	L	L	F	L	K	E	G	G	L	C	A	A	L	K	E	E	C	C	---	46.2%	
MCFFV	---	L	Q	N	R	R	G	L	D	-	L	L	F	L	K	E	G	G	L	C	A	A	L	K	E	E	C	C	---	46.2%	
KMLV	---	L	Q	N	R	R	G	L	D	-	L	L	F	L	K	E	G	G	L	C	A	A	L	K	E	E	C	C	---	46.2%	
HTLV-II	---	A	Q	N	R	R	G	L	D	-	L	L	F	W	E	Q	G	G	L	C	K	A	I	Q	E	Q	C	C	---	42.3%	
FeLV	---	L	Q	N	R	R	G	L	D	-	L	L	F	L	Q	E	G	G	L	C	A	A	L	K	E	E	C	C	---	42.3%	
FeSV	---	L	Q	N	R	R	G	L	D	-	I	L	F	L	Q	G	G	G	L	C	A	A	L	K	E	E	C	C	---	42.3%	
hom	---	N	R		D		L		G		C									C	C									30.8%	
sim	---	N	R		+	D	+	+	L		G									C	C									50.0%	
cons	---	l	q	n	r	r	g	l	d	-	l	l	-	-	-	-	-	-	-	g	g	-	c	-	a	l	-	e	-	c	c

FIG. 5. The conserved domain common to filoviral and retroviral envelope proteins. The conserved domains are 26 amino acid residues long and are located at the C-terminal ends of the envelope proteins. Percentages of homology to the MBG GP are shown. The GenBank-EMBL accession codes for the envelope protein sequences are given in Materials and Methods. SRV-1, simian AIDS retrovirus; MPMV, Mason-Pfizer monkey virus; RSV, Rous sarcoma virus; BAEV, baboon endogenous virus; ASV, avian sarcoma virus; ARV, avian reticuloendotheliosis virus; HTLV-I, human T-cell leukemia virus type I; GALV, gibbon ape leukemia virus; MoLV, Moloney murine leukemia virus; MCFFV, mink cell focus-forming virus; KMLV, Kirsten murine leukemia virus; HTLV-II, human T-cell leukemia virus type II; FeLV, feline leukemia virus; FeSV, feline sarcoma virus; hom, homologous residues; sim, similar residues; cons, consensus sequence. Dashes indicate gaps required for optimal alignment.

conserved transcription signals, as has been found with all MBG genes (7) and the NP-encoding gene of EBO (28).

The only transcript of gene 4 detected so far is a monocistronic mRNA with approximately the size of 28S rRNA. It migrates in denaturing agarose gels in a manner similar to that of the MBG NP mRNA transcript (7, 29). The transcript is polyadenylated, as could be shown for all MBG-specific mRNA transcripts by binding to oligo(dT) (7, 17). The coding region of the GP-encoding gene has a length of 2,045 nucleotides (positions 5939 to 7984) (Fig. 1 and 2). There is a second ORF in the messenger sense that extends between positions 6558 and 6896. Although transcription or translation products that correspond to this region have not been detected, it will be interesting to see whether the virus has developed mechanisms to utilize this genetic information. The noncoding regions at the 3' and 5' ends of gene 4 (118 and 681 nucleotides, respectively) are relatively long compared with those of most other nonsegmented negative-strand RNA viruses but resemble in length the untranslated regions of all other MBG mRNA transcripts (7) and of the NP mRNA of EBO (28).

The MBG GP is 681 amino acids long (Fig. 1, 2, and 4), and the unglycosylated polypeptide has a calculated M_r of 74,797. The sequence contains 22 Asn-X-Ser(Thr) motifs which are potential N-glycosylation sites. The Asn-Pro-Thr motif at position 389 is not a likely attachment site, because the proline residue in this specific position is known to interfere with glycosylation (1). Furthermore, there are overlapping motifs at positions 325 and 326 (Asn-Asn-Thr-Thr) and 344 and 345 (Asn-Asn-Thr-Ser), each of which is probably glycosylated at only one asparagine residue (39). It is therefore reasonable to assume that the GP has 19 N-glycosylation carbohydrate side chains that should account for a total of 55,000 to 60,000 daltons.

However, most of the oligosaccharides attached to the GP

represent O-glycans consisting of the core unit galactose- β (1-3)-*N*-acetylgalactosamine linked to serine or threonine residues (8, 11). Although evidence for a strict consensus sequence as established for N-glycans has not been obtained for O-glycosylation, one or several of the following features seem to be important: small clusters of hydroxyamino acids present in hydroxyamino acid-rich domains, proximity of a proline residue indicating turn or loop regions, and location adjacent to the anchor region in membrane proteins (2, 22, 35). Indeed, several clusters of hydroxyamino acids have been found in the middle part of the GP sequence (positions 170 to 500) which, in general, is remarkably rich in hydroxyamino acids (28.8% serine and threonine residues). Proline residues are associated with most of these hydroxyamino acid clusters. It is therefore reasonable to assume that the O-glycosidically linked carbohydrate chains are attached to the serine and/or threonine residues in these clusters. Previous studies (11) revealed a high amount of neutral mucin-type O-glycans with an average of four sugar residues (M_r , 1,200). On the basis of molecular weights of 170,000, 60,000, and 75,000 for the mature GP (8), all N-glycans, and the unglycosylated form of the GP, respectively, the M_r of O-linked oligosaccharide chains amounts to approximately 35,000. Thus, the MBG GP contains about 19 N-linked and 30 O-linked carbohydrate chains. This confirms previous results showing that 55 mol% of the oligosaccharides belong to O-glycans (11). These data, together with previous results (8, 11), demonstrate that carbohydrate structures account for more than 50% of the M_r of the MBG GP.

Comparisons of signal peptide sequences of different secretory and membrane proteins do not reveal a specific consensus sequence. All signal peptides, however, share a central core of uncharged hydrophobic amino acid residues and often show a positively charged residue (Lys or Arg) at the N-terminal end (41). These conditions are also met by the signal sequence at the amino terminus of the GP. Within the signal which has the sequence Met-Lys-Thr-Thr-Cys-Phe-Leu-Ile-Ser-Leu-Ile-Leu-Ile-Gln-Gly-Thr-Lys-Asn, there is a strong hydrophobic core at positions 6 to 13 and a positively charged lysine at position 2 at the amino-terminal end. These data further strengthen the amino acid sequence data in showing that the amino terminus of the GP serves as the signal peptide for translocation into the lumen of the endoplasmic reticulum.

Cleavage of the signal peptide is a cotranslational event catalyzed by a cellular signal peptidase located at the luminal site of the endoplasmic reticulum membrane (42). The specificity of the cleavage reaction is high despite a limited degree of conservation found among cleavage site sequences of different signal peptides (37). Nevertheless, it was possible to establish rules for certain amino acids preferred at specific positions around the cleavage site (23, 37, 38). With Asn, Thr, Lys, and Gly in positions -1, -3, -2, and -4, respectively, the absence of Pro in the region from -3 to +1 and the C terminus of the signal peptide hydrophobic core at -6, the cleavage site of the MBG GP follows most of these rules. The asparagine which is usually only a weak residue for position -1 can, if forced, fit into the -1 site (14).

Most viral glycoproteins are class I membrane proteins attached to the lipid bilayer by a single transmembrane region. Comparisons of several anchor regions did not result in a specific amino acid sequence, but anchors usually consists of 20 to 27 uncharged, hydrophobic amino acid residues. Besides the hydrophobic region at the N-terminal end, the GP has a sequence of 30 uncharged, primarily

hydrophobic amino acids located close to its C terminus (positions 644 to 673; Fig. 2 and 4). The hydrophobic region is followed by a short hydrophilic C-terminal tail of eight amino acid residues (positions 674 to 681; Fig. 2 and 4) which seems to be the cytoplasmic domain of the GP. The two positively charged residues, arginine 674 and lysine 678, probably work as a stop transmembrane signal, as in other viral membrane glycoproteins. These data, taken together, indicate that the MBG GP is a type I transmembrane protein.

Nucleotide sequence analyses of the NP-encoding and the L-protein-encoding gene of MBG and EBO suggested a common evolutionary history for all nonsegmented negative-stranded RNA viruses and showed that filoviruses are more closely related to paramyxoviruses than to rhabdoviruses (21, 28, 29). On the basis of these data, respiratory syncytial virus showed the highest similarity to the filoviruses. In contrast, there is no significant homology between the MBG GP and either one of the glycoproteins of respiratory syncytial virus. The G protein of respiratory syncytial virus superficially resembles the MBG GP in also containing both N-glycans and O-glycans, but its total carbohydrate content is significantly lower than that of the GP (40). In fact, together with gp160 of human immunodeficiency virus type 1 (10) and gp350 of Epstein-Barr virus (33) the GP is among the most highly glycosylated viral glycoproteins. The unusually high content of hydroxyamino acids and proline residues and the large amount of O-glycans are features also found in various mucinous proteins (18). It is generally assumed that carbohydrates may have a masking effect on functionally important sites of a surface protein, such as antigenic epitopes. It will therefore be interesting to see whether the large carbohydrate complement of the GP provides an explanation for some biological peculiarities of MBG, e.g., its inefficiency in inducing neutralizing antibodies.

Little is known about the pathogenetic mechanisms that underlie hemorrhagic fevers induced by filoviruses. Infection of human endothelial cells, a suitable target for MBG, results in cell destruction, thus interfering with the functions of the endothelium as a barrier between the blood and the surrounding tissues (31). Furthermore, there is evidence that infection of monkeys with EBO produces a marked depression of the immune response which is not directly associated with virus replication but seems to depend on humoral factors (26). It was therefore of particular interest that a sequence of 26 amino acids was discovered recently at the carboxy terminus of the EBO GP that showed homology with regions in the envelope glycoproteins of retroviruses that are believed to be responsible for immunosuppressive properties of these viruses (5, 36). Our data show that this domain is also present in the MBG GP and that it contains two highly conserved motifs, (i) a hexapeptide at the amino terminus (Fig. 5; N-R-r-g-l-D) which may function as the active core of the domain and (ii) three cysteine residues at the carboxy terminus which may be important for correct folding of the domain. It remains to be seen, however, whether MBG indeed has such an immunosuppressive function and whether the conserved domain plays a role in it.

ACKNOWLEDGMENTS

The excellent technical assistance of Friederike Mahner and Hans-Günter Welker is greatly acknowledged. We thank Beate Siegmann-Langner for preparing the schematic illustration, Bernd Becker for photographic work, and Rüttger Ebendt for assistance with the computer analyses. We greatly appreciate Anthony Sanchez for helpful discussion.

This work was supported by the Deutsche Forschungsgemeinschaft (SFB 286, Teilprojekt A6, and SFB 272, Teilprojekt Z2) and by the Fonds der Chemischen Industrie. Elke Mühlberger was a recipient of a research fellowship of the FAZIT-Stiftung.

REFERENCES

1. Bause, E. 1983. Structural requirements of N-glycosylation of proteins. *Biochem. J.* **209**:331-336.
2. Blochberger, T. C., J. M. Sabatine, Y. C. Lee, and R. P. Hughey. 1989. O-linked glycosylation of rat renal γ -glutamyltranspeptidase adjacent to its membrane anchor. *J. Biol. Chem.* **264**:20718-20722.
3. Bressan, G. M., and K. K. Stanley. 1987. pUEX, a bacterial expression vector related to pEX with universal host specificity. *Nucleic Acids Res.* **15**:10056.
4. Cattaneo, R. 1991. Different types of messenger RNA editing. *Annu. Rev. Genet.* **26**:71-88.
5. Cianciolo, G. J., T. D. Copeland, S. Oroszlan, and R. Snyderman. 1985. Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. *Science* **230**:453-455.
6. Elliott, L. H., M. P. Kiley, and J. B. McCormick. 1985. Descriptive analysis of Ebola virus proteins. *Virology* **163**:169-176.
7. Feldmann, H., E. Mühlberger, A. Randolph, C. Will, M. P. Kiley, A. Sanchez, and H.-D. Klenk. 1992. Marburg virus, a filovirus: messenger RNAs, gene order, and regulatory elements of the replication cycle. *Virus Res.* **24**:1-19.
8. Feldmann, H., C. Will, M. Schikore, W. Slenczka, and H.-D. Klenk. 1991. Glycosylation and oligomerization of the spike protein of Marburg virus. *Virology* **182**:353-356.
9. Gear, J. S. S., G. A. Cassel, A. J. Gear, B. Trappler, L. Clausen, A. M. Meyers, M. C. Kew, T. H. Bothwell, R. Sher, G. B. Miller, J. Schneider, H. J. Koornhoff, E. D. Comperts, M. Isaäcson, and J. H. S. Gear. 1975. Outbreak of Marburg virus disease in Johannesburg. *Br. Med. J.* **4**:489-493.
10. Geyer, H., C. Holschbach, G. Hunsmann, and J. Schneider. 1988. Carbohydrate of human immunodeficiency virus. Structures of oligosaccharides linked to the envelope glycoprotein 120. *J. Biol. Chem.* **263**:11760-11767.
11. Geyer, H., C. Will, H. Feldmann, H.-D. Klenk, and R. Geyer. 1992. Carbohydrate structure of Marburg virus glycoprotein. *Glycobiology* **2**:299-312.
12. Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* **25**:263-269.
13. Hewick, R. M., M. W. Hunkapillar, L. E. Hood, and W. J. Dreyer. 1981. A gas-liquid solid phase peptide and protein sequencer. *J. Biol. Chem.* **256**:7990-7997.
14. Hotin, G., and I. Boime. 1981. Miscalculation at the presequence of rat prolactin synthesized in pituitary cells incubated with a threonine analog. *Cell* **24**:453-461.
15. Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus. 1988. Signals for ribosomal frameshifting in the Rous sarcoma gag-pol region. *Cell* **55**:447-458.
16. Kiley, M. P., E. T. W. Bowen, G. A. Eddy, M. Isaäcson, K. M. Johnson, J. B. McCormick, F. A. Murphy, F. R. Pattyn, D. Peters, O. W. Prozesky, O. R. L. Regnery, D. I. H. Simpson, W. Slenczka, P. Sureau, G. van der Groen, P. A. Webb, and H. Wulff. 1982. Filoviridae: a taxonomic home for Marburg and Ebola viruses? *Intervirology* **18**:24-32.
17. Kiley, M. P., N. J. Cox, L. H. Elliott, A. Sanchez, R. DeFries, M. J. Buchmeier, D. D. Richman, and J. B. McCormick. 1988. Physicochemical properties of Marburg virus: evidence for three distinct virus strains and their relationship to Ebola virus. *J. Gen. Virol.* **69**:1975-1967.
18. Kornfeld, R., and S. Kornfeld. 1980. Structure of glycoproteins and their oligosaccharide units, p. 1-34. *In* W. J. Lennarz (ed.), *The biochemistry of glycoproteins and proteoglycans*. Plenum Press, New York.
19. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**:283-292.
20. Martini, G. A., and R. Siebert. 1971. Marburg virus disease, p. 1-230. Springer Verlag, New York.

21. Mühlberger, E., A. Sanchez, A. Randolph, C. Will, M. P. Kiley, H.-D. Klenk, and H. Feldmann. 1992. The nucleotide sequence of the L gene of Marburg virus, a filovirus: homologies with paramyxoviruses and rhabdoviruses. *Virology* **187**:534-547.
22. Niemann, H., R. Geyer, H.-D. Klenk, D. Linder, S. Stirn, and M. Wirth. 1984. The carbohydrates of mouse hepatitis virus (MHV) A59: structures of the O-glycosidically linked oligosaccharides of glycoprotein E1. *EMBO J.* **3**:665-670.
23. Perlman, D., and H. O. Halvorson. 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J. Mol. Biol.* **167**:391-409.
24. Pringle, C. R. 1991. The genetics of paramyxoviruses, p. 1-40. *In* D. W. Kingsbury (ed.), *Paramyxoviruses*. Plenum Press, New York.
25. Regnery, R. L., K. M. Johnson, and M. P. Kiley. 1981. Marburg and Ebola viruses: possible members of a new group of negative strand viruses, p. 971-977. *In* D. H. L. Bishop and R. W. Compans (ed.), *The replication of negative strand viruses*. Elsevier/North Holland, New York.
26. Ryabchikova, E. I., V. K. Tkachev, L. V. Kolesnikova, and J. N. Rassadkin. 1991. International Conference on Medical Biotechnology, Immunization, and AIDS, Leningrad, Russia. (Abstract in Russian.)
27. Sanchez, A., and M. P. Kiley. 1987. Identification and analysis of Ebola virus messenger RNA. *Virology* **157**:414-420.
28. Sanchez, A., M. P. Kiley, B. P. Holloway, J. B. McCormick, and D. D. Auperin. 1989. The nucleoprotein gene of Ebola virus: cloning, sequencing, and in vitro expression. *Virology* **170**:81-91.
29. Sanchez, A., M. P. Kiley, H.-D. Klenk, and H. Feldmann. 1992. Sequence analysis of the Marburg virus nucleoprotein gene: comparison to Ebola virus and other nonsegmented negative-strand RNA viruses: *J. Gen. Virol.* **73**:347-357.
30. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
31. Schnittler, H. J., F. Mahner, H. Wiener, D. Drenkhahn, H.-D. Klenk, and H. Feldmann. Replication of Marburg virus in human endothelial cells: a possible mechanism for the development of viral hemorrhagic disease. *J. Clin. Invest.*, in press.
32. Schuy, W., C. Will, K. Kuroda, C. Scholtissek, W. Garten, and H.-D. Klenk. 1986. Mutations blocking the transport of the influenza virus hemagglutinin between the rough endoplasmic reticulum and the Golgi apparatus. *EMBO J.* **5**:2831-2836.
33. Serafini-Cessi, F., N. Malagolini, M. Nanni, F. Dall'Olio, G. Campadelli-Fiume, J. Tanner, and E. Kieff. 1989. Characterization of N- and O-linked oligosaccharides of glycoprotein 350 from Epstein-Barr virus. *Virology* **170**:1-10.
34. Smith, D. H., B. K. Johnson, M. Isaacson, R. Swanepoel, K. M. Johnson, M. P. Kiley, A. Bagshawe, T. Siongok, and W. K. Keruga. 1982. Marburg-virus disease in Kenya. *Lancet* **i**:816-820.
35. Tomita, M., H. Furthmayr, and V. T. Marchesi. 1978. Primary structure of human erythrocyte glycophorin A. Isolation and characterization of peptides and complete amino acid sequence. *Biochemistry* **17**:4756-4770.
36. Volchkov, V. E., V. M. Blinov, and S. V. Netesov. 1992. The envelope glycoprotein of Ebola virus contains an immunosuppressive-like domain similar to oncogenic retroviruses. *FEBS Lett.* **305**:181-184.
37. von Heijne, G. 1983. Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* **133**:17-21.
38. von Heijne, G. 1984. How signal sequences maintain cleavage specificity. *J. Mol. Biol.* **73**:243-251.
39. Waterfield, M. D., M. J. Gething, G. Scrace, and J. J. Skehel. 1980. The carbohydrate side chains and disulfide bonds of the hemagglutinin of the influenza virus A/Japan 305/57(H2N1), p. 11-20. *In* G. Laver and G. Air (ed.), *Structure and variation in influenza virus*. Elsevier/North Holland, New York.
40. Wertz, G. W., P. L. Collins, Y. Huang, C. Gruber, S. Levine, and L. A. Ball. 1985. Nucleotide sequence of the G protein gene of human respiratory syncytial virus reveals an unusual type of viral membrane protein. *Proc. Natl. Acad. Sci. USA* **82**:4075-4079.
41. Wiley, D. C. 1990. Viral membranes, p. 45-67. *In* B. N. Fields and D. M. Knipe (ed.), *Fundamental virology*. Raven Press, New York.
42. Zwizinski, C., and W. Wickner. 1980. Purification and characterization of leader (signal) peptidase from *Escherichia coli*. *J. Biol. Chem.* **255**:7973-7977.