Volume 305, number 3, 181–184 FEBS 11222 July 1992

The envelope glycoprotein of Ebola virus contains an immunosuppressive-like domain similar to oncogenic retroviruses

Viktor E. Volchkov, Vladimir M. Blinov and Sergey V. Netesov

Institute of Molecular Biology, Koltsovo, 633159 Novosibirsk Region, Russia

Received 7 May 1992

Genomic RNA of a Zaire strain of Ebola virus was cloned, and cDNA inserts specific for the glycoprotein gene were isolated and sequenced. The determined sequence has only one open reading frame encoding 318 amino acids and is part of ORF-4 on the plus RNA strand. The putative transcriptional stop site (3' AAUUCUUUUU 5') and the transcriptional start site (3' AACUACUUCUAAUU.. 5') were identified. Computer-assisted comparison of the amino acid sequence of the C-terminal part of protein encoded by ORF-4 of Ebola virus with sequences of the proteins present in the SWISSPROT and EMBL banks revealed significant homology with the 'immunosuppressive domain' of the p15E envelope proteins of various oncogenic retroviruses. The possible role of such a homology is discussed.

Ebola virus; Immunosuppressive domain; cDNA; Amino acid sequence

1. INTRODUCTION

Ebola virus is an enveloped, negative-stranded RNA virus. This virus belongs to the family Filoviridae and has a distinctive morphology similar only to that of Marburg virus, but apparently has no antigenic characteristics of Marburg virus [1,2]. Its genome is a non-segmented RNA strand with an approximate molecular weight of $4 \cdot 10^6$. SDS-PAGE profiles demonstrate the presence of seven structural proteins. These are the major nucleoprotein NP (104 K), glycoprotein GP (125 K), L protein (180 K), and four proteins: VP40 (40 K), VP35 (35 K), VP30 (30 K) and VP24 (24 K). All the proteins are encoded by monocistronic mRNA transcripts complementary to virion RNA [3,4]. GP is the major protein of the surface spikes, and only this 125 K viral protein is glycosylated [3,5].

To better understand the nature of this exclusively pathogenic agent of a human infectious disease, it is necessary to study the organization of Ebola virus genome. In this paper we present cloning and sequencing data of the fragment of virion RNA encoding the C-terminus of the G-protein. In addition we compare this region with sequences from EMBL and SWISS-PROT banks and reveal that it has homology with the 'immunosuppressive domain' of oncogenic retroviruses.

Abbreviations: ARV, avian reticuloendotheliosis virus (retrovirus); FeLV, feline leukemia virus (retrovirus); RSV, Rous sarcoma virus (retrovirus); ASV, avian sarcoma virus (retrovirus); BAEV, baboon endogenous virus (retrovirus); HTLV1C, human T-cell leukemia virus type I; HTLV2, human T-cell leukemia virus type II (retrovirus).

Correspondence address: V.E. Volchkov, Institute of Molecular Biology, Koltsovo, 633159 Novosibirsk Region, Russia. Fax: (7) (383) (2) 642 974.

2. MATERIALS AND METHODS

The Mayinga strain of Ebola virus (a Zaire subtype) had been obtained from the Byelorussian Institute of Microbiology and Epidemiology (Minsk, USSR) and was onze passaged in *Macaca* rhesus before use. The virus was cultured in Vero cells and purified from the tissue culture liquid as described previously [3–5]. The virus was resuspended in TNE and used for RNA extraction. Extraction of virion RNA was carried out as described in [6]. The first-strand cDNA synthesis on viral RNA templates was primed with a random primer using RT. RNA–cDNA hydrids were poly(dC)-tailed and cloned in *Pst*I digested poly(dG)-tailesdvector plasmid pBR322. Recombinant plasmids with the longest prolonged cDNA inserts were identified by in situ colony hybridization. DNA sequences were determined according to Maxam and Gilbert [7]. Homology search through SWISSPROT or the EMBL bank was performed using the QUICK program of the GENEBEE package.

3. RESULTS AND DISCUSSION

Standard gene-engineering techniques were used to obtain a wide collection of recombinant clones containing inserts of cDNA specific for Ebola virus. Seven partly overlapping inserts of cDNA containing more than 90% of virion RNA sequence were chosen for nucleotide sequencing.

Here we present the partial nucleotide sequence of 3'-end of the ORF-4-encoding gene (1,194 bp) from the plus RNA-strand (Fig. 1). This nucleotide sequence was determined by sequencing the inserts of pEb137 and pEb102 plasmids (more than 95% of the DNA sequence was determined for both strands). The full-length sequence of the Ebola virus genome will soon be published elsewhere.

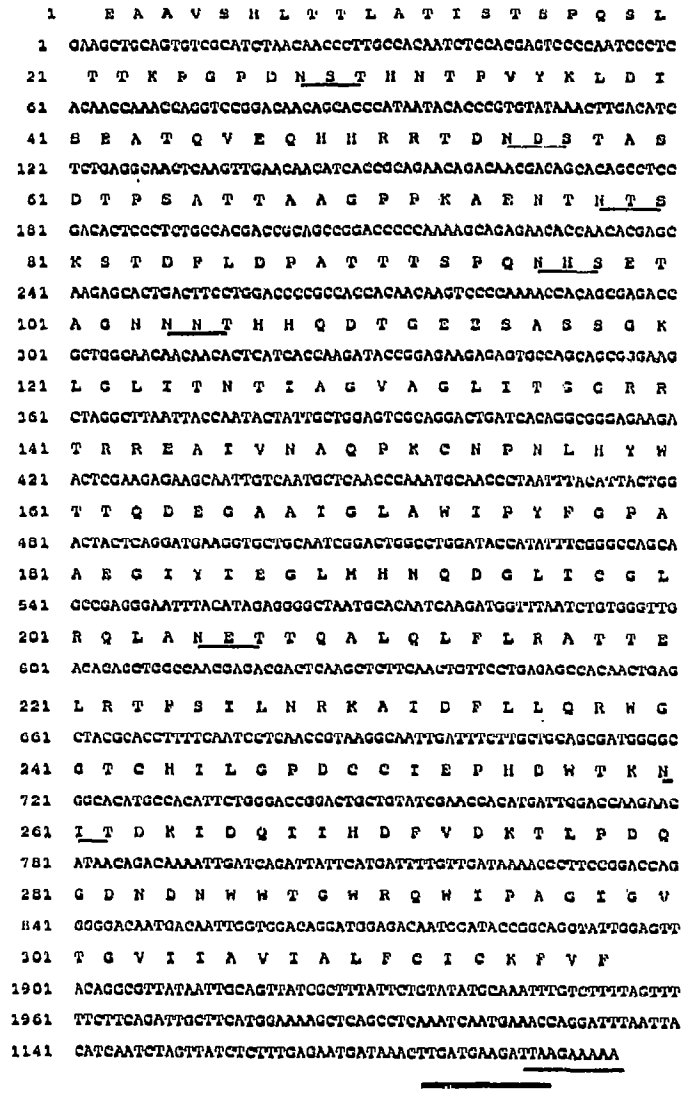

Fig. 1. Nucleotide sequence of 1,194 bp of the part of viral inserts in pEB137 and pEB102. The sequence represented is the positive anti-genome strand in the DNA form. The possible translation product is indicated. The one-letter code for abbreviating amino acids is used. Potential glycosylation sites are denoted by single lines under each of the tripeptide sequences. The stop and start transcription signals are indicated by the single and double underlines, respectively.

The sequence determined showed that there was only one open reading frame, which encoded 318 amino acids. The transcriptional stop site (5′ ..UUAAGAAAAA 3′) on the 3′-end of the sequence was predicted according to its homology with the NP gene of Ebola virus [8]. The putative transcriptional start site for mRNA-5 (5′ UUGAUGAAGAUUAA ..3′) was overlapped by the transcriptional stop site of mRNA-4 and was identified on the basis of homology with the corresponding site of the gene [8]. Similar transcriptional signals were found for all the seven prolonged ORFs of the Ebola plus RNA strand (data not shown).

We supposed this sequence to constitute part of the GP gene. Recently it was demonstrated that the GP protein of a related Ebola virus – Marburg virus – was encoded by ORF-4 of the plus RNA strand [9]. Moreover, the GP protein of Ebola virus is known to have a large molecular weight (125 K) and is the only viral glycoprotein [3,5]. The molecular weight of the protein encoded by the predicted ORF-4 is about 75 kDa and contains 12 potential asparagine-linked glycosylation sites, 7 of them being indicated on the sequence shown in Fig. 1.

There are data indicating a marked depression of the immune response observed in Ebola infection of monkeys which is not directly associated with virus reproduction, but obviously depends on 'humoral' factors [10]. One of the proteins of Ebola virus is supposed by us to have an immunosuppressive-like domain which plays a crucial role in the pathogenesis of this infection.

The deduced amino acid sequences of all prolonged ORFs found on the plus RNA strand of Ebola virus were compared with sequences of proteins with immunosuppressive properties present in the SWISS-PROT and EMBL banks. This analysis indicated the occurrence of a significant level of homology with the p15E proteins of the various oncogenic retroviruses.

Retroviruses are known to cause a large number of animal cancers often associated with immunosuppression [11–14]. There is evidence that certain retrovirus virion proteins synthesized by infected cells may contribute to the immunosuppression which has been linked to a conserved segment of 18 amino acids, the so-called 'immunosuppressive peptide', located within the transmembrane component of the envelope protein. This sequence results in immunodeficiency of the host organism by blocking the induction of lymphocytes in response to antigens or mitogens. The immuno-suppressive propgerties of the envelope protein of Avian reticuloendotheliosis virus (ARV), notably the inhibition of the blastogenic response of lymphocytes to Con-A by induction of a T-suppressor cell population, are well described [15]. A synthetic peptide representing 14 amino acids of Feline leukemia virus (FeLV) p15E is known to inhibit the proliferation of lymphoid cells in vitro [16]. Furthermore, the disrupted Baboon endogenous virus (BAEV), also sharing the p15E sequence, is capable of blocking CnA-induced blastogenesis of human lymphocytes [17].

We compared the amino acid sequences of the envelope proteins of ARV, RSV, ASV, BAEV, HTLV1C and HTLV2 [18–23] with the sequence of the C-terminal part of the protein encoded by ORF-4 of Ebola virus. As shown in Fig. 2, the homologous regions between Ebola virus and retroviruses involve 160 residues, including six cysteine residues. Over this region the envelope polyproteins of retroviruses and Ebola virus ex-

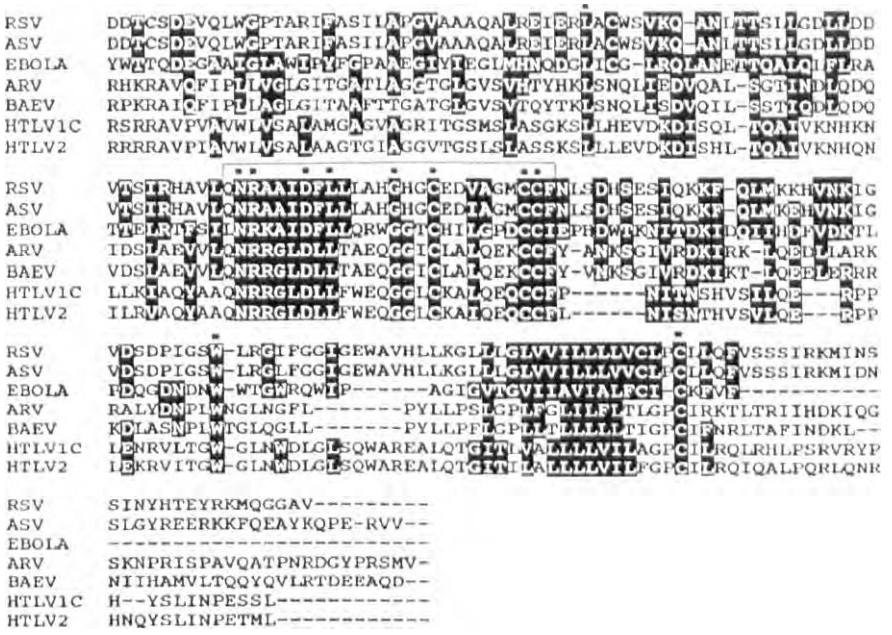

Fig. 2. Amino acid sequence alignment of the C-terminal ORF-4 of Ebola virus and p15E-related proteins of Rous sarcoma virus (RSV), Avian sarcoma virus (ASV), ARV, BAEV, Human T-cell leukemia virus type I (HTLV1C) and HTLV2. Dots indicate amino acid residues identical or belonging to the same group of chemical similarity as Ebola virus. The amino acid families are separated by us as follows: (P,G,S,T,A); (F,Y,W,I,L,M,V); (D,E,N,Q); (K,R,H); (C). Gaps are introduced to maximize homology. The one-letter code is used to abbreviate amino acids. Residues identical for all proteins are marked by asterisks. The immunosuppressive domain is denoted by single lines.

hibit 24–44 identities, but this optimal alignment requires the insertion of 10–12 gaps. In spite of the amino acid sequence divergence, these proteins exhibit a number of structural similarities. Firstly, as shown in Fig. 2, the immunosuppressive domain of retroviruses and the similar domain of Ebola virus (residues 227–252) have high homology (38–46%). Secondly, the greater part of the amino acid substitutions is conserved, and moreover, retroviruses and Ebola virus share potential sites of glycosylation in these regions. Three cysteine residues in this region are also conserved, which suggests similar patterns of disulfide bonding.

Hypothetical models of the association of p15E-related proteins of retroviruses and the C-terminus of the ORF-4 protein (pGPC) with the membrane are similar (Fig. 3). This is confirmed by calculating the hydrophobicity index. In this model the main part of p15E protein and pGPC is located outside the lipid portion of the membrane and anchored to the membrane via a hydrophobic sequence near the carboxyl terminus.

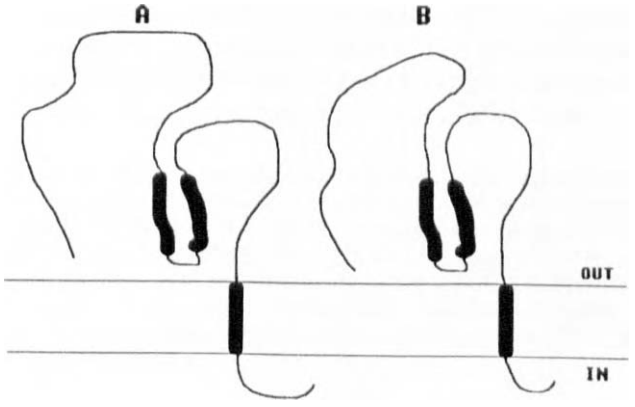

Fig. 3. The model of a possible structure of the envelope polyproteins such as that of the retroviruses [24] (A), and the C-terminus of the glycoprotein of Ebola virus (B) to the hydrophobic part of the membrane. Three hydrophobic regions are denoted by the black boxes.

Therefore we offer as a working hypothesis that the presence of such a well-conserved region occurring in the envelope proteins of retroviruses and Ebola virus may be accounted for by recombination between the viral RNAs. The evolutionary implications of this hypothesis is discussed but RNA recombination has been reported for different viruses [24].

For the similarity of the envelope proteins of retroviruses and the protein encoded by ORF-4 of Ebola virus, it is possible that a similar mechanism might be functional in the pathogenicity of these viruses. It remains to be determined whether the immunosuppressive domain is located on the outside membrane of virus particles and virus-infected cells and wether it plays a role in pathogenicity of Ebola virus infection.

REFERENCES

[1] Johnson, K.M., Webb, P.A., Lange, J.V. and Murphy, F.A. (1977) Lancet 1, 569–571.
[2] Bowen, E.T.W., Platt, G.S., Lloyd, G., Baskerville, A., Harris, W.J. and Vella, E.E (1977) Lancet 1, 571–573.
[3] Elliott, L.H., Kiley, M.P., McCormick, J.B. (1985) Virology 147, 169–176.

[4] Sanchez, A. and Kiley, M.P. (1987) Virology 157, 414–420.
[5] Kiley, M.P., Bowen, E.T.W., Eddy, G.A., Isaacson, M., Johnson, K.M., McCormick, J.B., Murphy, F.A., Pattyn, S.R., Peters, D., Prosersky, O.W., Regnery, R.L., Simpson, D.I.H., Slenczka, W., Sureau, P., Van Der Groen, G., Webb, P.A. and Wulff, H. (1982) Intervirology 18, 24–32.
[6] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
[7] Maxam, A. and Gilbert, W. (1980) in: Methods in Enzymology (L. Grossman and K. Moldave Eds.) vol. 65, Academic Press, New York, pp. 499–560.
[8] Sanchez, A., Kiley, M.P., Holloway, B.P., McCormick, J.B. and Auperin, D.D. (1989) Virology 170, 81–91.
[9] Feldmann, H., Muhlberger, E., Randolf, A., Will, C., Sanchez, A. and Klenk, H.-D. (1991) 8th International Conference on Negative Strand Viruses, Charleston, SC, USA.
[10] Ryabchikova, E.I., Tkachev, V.K., Kolesnikova, L.V. and Rassadkin, Ju. N. (1991) in: International Conference on Medical Biotechnology, Immunization and AIDS, Leningrad, USSR.
[11] Sonigo, P., Barker, C., Hunter, E. and Wain-Hobson, S. (1986) Cell 45, 375–385.
[12] Gallo, R.C. and Wong-Staal, F. (1980) in: Viral oncology, (Klein, G. ed.) Elsevier, Amsterdam, New York.
[13] Perryman, L.E., Hoover, E.A. and Yohn, D.S. (1972) JNCI 43, 1357–1363.
[14] Mathes, L.E., Olsen, R.G., Hebebrand, L.C., Hoover, E.A., Schaller, J.P., Adams, P.W. and Nichols, W.S. (1979) Cancer Res. 39, 950–955.
[15] Rub, B.J., Spence, J.L., Haelzer, J.D., Lewis, R.B., Carpenter, C.R., Rubin, A.S. and Bose, H.R. (1979) J. Immunol. 123, 1362–1370.
[16] Cianciolo, G.J., Copeland, T.D., Oroszlan, S. and Snyderman, R. (1985) Science 230, 453–455.
[17] Denner, J., Wunderlich, V. and Bierwolf, D. (1980) Acta Biol. Med. Ger. 39, K19–K26.
[18] Wilhelmsen, K.C., Eggleton, K. and Temin, H. (1984) J. Virol. 52, 172–182.
[19] Schwartz, D.E., Tizard, R. and Gilbert, W. (1983) Cell 32, 853–869.
[20] Neckameyer, W.S., Wang, L.-H. (1985) J. Virol. 53, 879–884.
[21] Kato, S., Matsuo, K., Nishimura, N., Takahashi, N. and Takano, T. (1987) Jpn. J. Genet. 62, 127–137.
[22] Malik, K.T.A., Even, J. and Karpas, A. (1988) J. Gen. Virol. 69, 1695–1710.
[23] Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D.W., Chen, I.S.Y., Miwa, M. and Sugimura, T. (1985) Proc. Natl. Acad. Sci. USA 82, 3101–3105.
[24] Romanova, L.I., Blinov, V.M., Tolskaya, E.A., Viktorova, E.G., Kolesnikova, M.S., Guseva, E.A. and Agol, V.I. (1986) Virology 155, 202–213.