

## Comprehensive Analysis of Ebola Virus GP1 in Viral Entry

Balaji Manicassamy,<sup>†</sup> Jizhen Wang,<sup>†</sup> Haiqing Jiang, and Lijun Rong\*

*Department of Microbiology and Immunology, College of Medicine, University of Illinois at Chicago, Chicago, Illinois*

Received 27 July 2004/Accepted 19 November 2004

**Ebola virus infection is initiated by interactions between the viral glycoprotein GP1 and its cognate receptor(s), but little is known about the structure and function of GP1 in viral entry, partly due to the concern about safety when working with the live Ebola virus and the difficulty of manipulating the RNA genome of Ebola virus. In this study, we have used a human immunodeficiency virus-based pseudotyped virus as a surrogate system to dissect the role of Ebola virus GP1 in viral entry. Analysis of more than 100 deletion and amino acid substitution mutants of GP1 with respect to protein expression, processing, viral incorporation, and viral entry has allowed us to map the region of GP1 responsible for viral entry to the N-terminal 150 residues. Furthermore, six amino acids in this region have been identified as critical residues for early events in Ebola virus entry, and among these, three are clustered and are implicated as part of a potential receptor-binding pocket. In addition, substitutions of some 30 residues in GP1 are shown to adversely affect GP1 expression, processing, and viral incorporation, suggesting that these residues are involved in the proper folding and/or overall conformation of GP. Sequence comparison of the GP1 proteins suggests that the majority of the critical residues for GP folding and viral entry identified in Ebola virus GP1 are conserved in Marburg virus. These results provide information for elucidating the structural and functional roles of the filoviral glycoproteins and for developing potential therapeutics to block viral entry.**

Ebola viruses are enveloped, single-stranded, negative-sense RNA viruses. Together with Marburg virus, they constitute the filovirus family. There are four species of Ebola viruses, namely, Zaire, Sudan, Reston, and Ivory Coast. Ebola virus infection causes severe hemorrhagic fever in humans and non-human primates, with a mortality rate of up to 90% (12). The outbreaks are infrequent and unpredictable. Currently there is no vaccine or antiviral therapy against Ebola virus infection for humans. Moreover, the natural reservoir for the virus is still unknown, and hence it has been difficult to devise strategies to control the transmission of infection. Research on Ebola virus in tissue cultures and animal models has been hindered by the requirement for maximum-containment facilities (biosafety level 4).

The Ebola virus genome has seven genes, NP, VP35, VP40, GP, VP30, VP24, and L (12). The GP gene encodes two proteins, a secreted form called sGP and a transmembrane form called GP which is produced by RNA editing (13, 14, 26). GP is the envelope glycoprotein that is present on the virion surface and is involved in receptor binding and mediating viral entry (5). GP is synthesized as a single polypeptide of 676 amino acids in length (12). The N-terminal 32 amino acids are the signal peptide, which is cleaved immediately after translation. The cleaved GP undergoes N glycosylation in the endoplasmic reticulum to form pre-GP (~100 kDa), and the pre-GP is further processed in the Golgi to produce fully glycosylated GP (GP0) (7, 27). GP0 is cleaved by a cellular protease (furin convertase) into a surface subunit, GP1 (~130

kDa), and a transmembrane subunit, GP2 (~24 kDa). GP1 is associated with GP2, possibly through both noncovalent interactions and disulfide bonding (7, 27). Similar to the case for retroviruses and influenza virus, the native form of GP on virions exists as a homotrimer of GP1/GP2 heterodimer (15). The GP1 subunit is responsible for receptor binding, and GP2 mediates virus-cell membrane fusion and viral entry, likely in a pH-dependent manner (24, 32). Structural analysis of the core domain of Ebola virus GP2 indicates that the fusion-active conformation of GP2 is similar to that of the other class I membrane fusion proteins, suggesting a common membrane fusion mechanism (10, 29, 30).

It has been previously shown that transient expression of GP in tissue culture causes cytopathic effects such as cell rounding and detachment (18, 25, 33). This cytopathic effect has been mapped to a mucin-like region in GP1, which is heavily modified by O glycosylation (18, 33). However, deletion of this region has no effect on entry and cell tropism (7, 33). These findings were corroborated by studies that showed that the endothelial cell lining in blood vessel explants was damaged by GP expression (33). Further, GP expression can down-regulate the expression of several cell surface molecules, such as major histocompatibility complex class I and integrins (18, 25), which are involved in immune signaling and cell adhesion, respectively. Nevertheless, the structural and functional information on Ebola virus GP1 is very limited. This lack of understanding can largely be attributed to (i) the biosafety concern about working with the live viruses, (ii) the difficulty of genetically manipulating the Ebola virus RNA genome, and (iii) the technical obstacles in purifying large amounts of biologically active protein and determining the structure of GP. In addition, the receptor(s) for Ebola virus entry has not been identified, although several cell surface proteins (such as folate receptor  $\alpha$ ,

\* Corresponding author. Mailing address: Department of Microbiology and Immunology, College of Medicine, University of Illinois at Chicago, E829 MSB, 835 S. Wolcott Ave., Chicago, IL 60612. Phone: (312) 355-0203. Fax: (312) 996-6415. E-mail: lijun@uic.edu.

<sup>†</sup> B.M. and J.W. contributed equally to this work.

DC-SIGN, and L-SIGN) have been implicated in facilitating Ebola virus infection (1, 2, 9, 16, 17, 23).

In this study, we have used a human immunodeficiency virus (HIV)-based pseudotyped virus as a surrogate system to map the region of GP1 important for receptor recognition and to identify and characterize the specific residues in this region that are involved in receptor recognition and protein folding. This system allows easy genetic manipulation of GP, with the safety concerns alleviated. We have found that Ebola virus entry is determined predominantly by the N-terminal 150 residues of GP1. Furthermore, we have identified more than 30 residues in GP1 as potentially critical residues in receptor recognition or protein folding and structure. Of particular interest, we have identified a potential GP1-receptor interaction pocket which may be an ideal target for blocking Ebola virus entry. In addition, sequence alignment suggests that most of the critical residues identified in Ebola virus GP1 are also conserved in Marburg virus. Based on these data, we propose that the N-terminal 150 amino acids of GP1 of Ebola and Marburg viruses likely adopt similar conformations and bind to the same or related receptors. These findings may help in the development of effective entry inhibitors against filoviral infection.

#### MATERIALS AND METHODS

**Cell lines and antibodies.** 293T human embryonic kidney cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum with penicillin and streptomycin plus 300  $\mu$ g of Geneticin per ml. HeLa cells were maintained in the same medium without Geneticin. The mouse monoclonal antibody 12B5-1-1, which recognizes the GP1 of Ebola virus Zaire (EBOZ) GP, was kindly provided by Mary K. Hart (U.S. Army Medical Research Institute of Infectious Diseases) (31). The mouse anti-HIV p24 monoclonal antibody was obtained from the National Institutes of Health AIDS Research and Reference Reagent Program (19).

**Mutagenesis of the EBOZ GP gene.** The Ebola virus Zaire glycoprotein gene was synthesized by multiple rounds of overlapping PCR based on the EBOZ genome sequence (GenBank accession number L11365) (H. Jiang et al., unpublished data). The  $\Delta$ mucin and  $\Delta$ 1 to  $\Delta$ 13 mutants were constructed by PCR-directed mutagenesis. All alanine substitution mutations of the EBOZ GP gene were generated by site-directed mutagenesis with the Stratagene Quick-Change mutagenesis kit according to the supplier's protocols. All mutations were confirmed by DNA sequencing of flanking regions (on average, approximately 500 bp). The full length of GP gene was confirmed by DNA sequencing for those mutants that displayed defective phenotypes (Table 1).

**Pseudotyping.** Pseudotyped viruses were produced by cotransfecting DNA of wild-type (wt) GP or mutants with Env-deficient HIV vector (pNL4-3-Luc-R<sup>-</sup>-E<sup>-</sup>) (4, 6) into producer cells. In those experiments, 2  $\mu$ g of DNA of wt GP or mutants and 2  $\mu$ g of pNL4-3-Luc-R<sup>-</sup>-E<sup>-</sup> were used to transfect 293T cells (90% confluent) in six-well plates with Lipofectamine 2000 according to the protocol of the supplier (Invitrogen). The supernatants containing the pseudotyped viruses were collected twice (at 24 and 48 h posttransfection), combined, clarified from floating cells and cells debris by low-speed centrifugation, and filtered through a 0.45- $\mu$ m-pore-size filter (Nalgene). Then, 1 ml of supernatant was used to infect cells, and the rest was stored at  $-80^{\circ}\text{C}$  until use.

**Western blotting.** To evaluate EBOZ GP expression, the 293T producer cells were lysed in 0.2 ml of Triton X-100 lysis buffer (50 mM Tris-HCl [pH 7.5], 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, and a protease inhibitors cocktail consisting of 10  $\mu$ g of leupeptin per ml, 5  $\mu$ g of aprotinin per ml, and 2 mM phenylmethylsulfonyl fluoride) at 48 h after cotransfection. The protein samples were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a polyvinylidene difluoride membrane. The membrane was first incubated with anti-EBOZ GP1 monoclonal antibody 12B5-1-1 (1:5000 dilution) for 1 h and then probed with peroxidase-conjugated goat anti-mouse antiserum (Pierce) for 1 h. The bands were visualized by the chemiluminescence method according to the protocol of the supplier (Pierce).

**Detection of GP incorporation in pseudotyped viruses.** To evaluate the incorporation of wt GP protein or mutants into the pseudotyped viruses, 2 ml of

pseudotyped virus was layered onto a 3-ml cushion of 20% (wt/vol) sucrose in phosphate-buffered saline and centrifuged at 55,000 rpm for 30 min in a Beckman SW41 rotor. The pelleted pseudotyped viruses were lysed in 50  $\mu$ l of Triton X-100 lysis buffer, and a 25- $\mu$ l sample was subjected to SDS-PAGE. Expression of the EBOZ GP protein or mutants was detected by Western blotting as described above. A mouse anti-HIV p24 monoclonal antibody (1:5,000 dilution) was used as the primary antibody to detect the HIV p24 protein.

**Assay of infectivities of pseudotyped viruses.** Human 293T or HeLa cells ( $3 \times 10^5$ ) were seeded in six-well plates 1 day prior to infection. These targeted cells were incubated with 1 ml of the pseudotyped viruses for 5 h. The cells were lysed in 200  $\mu$ l of cell culture lysis reagent (Promega) at 48 h postinfection. The luciferase activity was measured with a luciferase assay kit (Promega) and an FB12 luminometer (Berthold detection system) according to the supplier's protocol. Each experiment was done in triplicate and repeated at least three times.

#### RESULTS

To study the mechanism of entry of Ebola viruses with the alleviation of safety concerns, we have established an HIV-based pseudotyping surrogate system using a synthetic GP gene of the Zaire species (Jiang et al., unpublished data). This system allows us to quickly analyze the effects of a large number of GP mutations on viral entry. In this study, each of the GP mutants was examined for protein expression or processing, incorporation into HIV particles, and the ability to mediate viral entry. Since defects in GP expression or processing and viral incorporation are manifestations of GP misfolding and structural alterations, we have been able to identify numerous residues in GP1 that appear to play crucial roles in protein folding and/or maintenance of the structural integrity of GP. Furthermore, we have identified several critical residues in GP1 that may be directly involved in receptor interaction.

**Deletions of GP residues between positions 266 and 476 do not greatly affect viral entry.** First, a series of deletion mutants to delineate the region of GP1 essential for Ebola virus entry were constructed and characterized (Fig. 1A). The C terminus of GP1 contains a mucin-like region, and it was previously shown that this region is not essential for viral entry (7, 33). To confirm this finding, a  $\Delta$ mucin GP mutant ( $\Delta$ 309-476) was constructed. The protein expression was examined by Western blotting of 293T cell lysates transiently transfected with  $\Delta$ mucin GP plasmid. A band with apparent molecular mass of 60 kDa was detected in  $\Delta$ mucin GP transfected cells (Fig. 1B, lane 1), consistent with the previous reports (7, 33). However, compared to that of wild-type GP1 (Fig. 1B, lane 2), the level of  $\Delta$ mucin GP1 detected was greatly reduced. We also could not detect any  $\Delta$ mucin GP1 in the supernatant (data not shown). To generate GP-HIV-pseudotyped Ebola virus, the GP plasmids and HIV vector were cotransfected into 293T cells as described in Materials and Methods. The level of incorporation of GP protein was examined by Western blotting after the virions were pelleted on a 20% sucrose cushion. It is interesting that although the level of  $\Delta$ mucin GP1 was barely detectable in the cell lysate (Fig. 1B, lane 4), the amount of  $\Delta$ mucin GP1 incorporated into the HIV particles (lane 7) was comparable to or higher than that of the wt GP1 (lane 8). It is not clear why we could barely detect  $\Delta$ mucin GP1 in the cell lysate.

To examine the ability of the  $\Delta$ mucin GP to mediate viral entry, HIV virions pseudotyped with either wt GP or the mutant GP were used to challenge 293T cells, which are susceptible to infection by the Ebola GP-pseudotyped viruses, and

TABLE 1. Summary of Ebola virus GP1 substitution mutants

GP1	Protein expression	Viral incorporation	Infectivity (%)	Class <sup>a</sup>	Note
wt	++++	++++	100		Full length sequenced
M1 (D47A/D49A/K50A)	++++	++++	17	IV	Full length sequenced
M2 (R54A/D55A/K56A)	++++	+++	1	III	Full length sequenced
M6 (K84A/R85A)	+	+	0	II	Full length sequenced
M10 (E103A/E106A)	+	+	0	II	Full length sequenced
M11 (E112A/K114A/K115A)	++++	+++	10	IV	Full length sequenced
M12 (D117A/E120A)	+++	++	17	III	Full length sequenced
M13 (D127A/R130A)	++++	++++	71	I	
M14 (R134A/R136A)	++++	++	59	III	
M15 (H139A/K140A)	+	+	1	II	Full length sequenced
M17 (H154A/K155A)	+	+	1	II	Full length sequenced
M18 (D163A/R164A)	++++	+++	58	I	
M21 (K190A/K191A)	++++	++++	44	I	
M22 (R200A/E201A)	++++	++++	77	I	
M23 (E207A/D208A)	++++	++++	75	I	
M25 (E235A/D237A)	++++	++++	54	I	
L43A	++++	+++	45	III	
D47A	++++	++++	53	I	
D49A	++++	++++	55	I	
K50A	++++	++++	89	I	
L51A	++++	+++	21	III	
R54A	++++	++++	50	I	
D55A	++++	++++	13	IV	Full length sequenced
D55E	++++	++++	111	I	
D55K	++++	++++	89	I	
K56A	++++	++++	46	I	
L57A	++++	+++	2	IV	Full length sequenced
L57I	++++	++++	7	IV	
L57F	++++	++++	5	IV	
L57K	++++	++++	5	IV	
Q62A	++++	++++	195	I	
L63A	++++	++++	14	IV	
L63I	++++	++++	78	I	
L63F	+++	++	2	II	
L63K	++	+	0	II	
M3 (R64A)	++++	++++	0	IV	Full length sequenced
R64K	++++	++++	120	I	
R64E	++++	+++	0	IV	
L68A	++	+++	3	II	
N69A	++++	++++	98	I	
L70A	++++	++++	95	I	
M4 (E74A)	+++	+	2	III	Full length sequenced
M5 (D78A)	++++	++++	115	I	
K84A	++++	++++	94	I	
R85A	+	+	0	II	Full length sequenced
W86A	+	++	8	II	
F88A	++++	++++	1	IV	Full length sequenced
F88I	++++	++++	104	I	
F88E	++++	++	0	III	
M7 (R89A)	+++	+++	59	I	
M8 (K95A)	++++	++++	8	IV	Full length sequenced, second mutation V48I
K95R	++++	+++	74	I	
K95E	++++	++	26	III	
Y99A	++	+	3	II	
M9 (E100A)	++++	++++	48	I	
E103A	+	+	0	II	Full length sequenced
W104A	++++	++++	92	I	
E106A	++++	++++	87	I	
N107A	++++	++++	79	I	
Y109A	+	+	9	II	
N110A	++++	++++	183	I	
L111A	++++	+	2	III	
E112A	++++	++++	67	I	
I113A	++++	++	25	III	
K114A	++++	++++	53	I	
K115A	++++	++++	56	I	

Continued on following page

TABLE 1—Continued

GP1	Protein expression	Viral incorporation	Infectivity (%)	Class <sup>a</sup>	Note
D117A	+++	+	17	III	
E120A	++++	++	35	III	
L122A	++++	++	4	III	Full length sequenced
F132A	++++	+	35	III	
Y137A	++++	+	75	III	Full length sequenced
H139A	+	+	2	II	Full length sequenced
K140A	++++	++++	74	I	
M16 (D150A)	++++	++++	135	I	
F153A	++++	++	31	III	
H154A	+	+	0	II	Full length sequenced
K155A	++++	++++	61	I	
F159A	++	++	1	II	
L161A	++	+	1	II	
Y162A	++	+	2	II	
L165A	++++	++++	55	I	
I170A	++++	++++	16	IV	Full length sequenced
I170F	++++	+	45	III	
I170E	++++	+	0	III	
Y171A	+	++	1	II	
M19 (R172A)	++++	++++	55	I	
F176A	+	+	0	II	
M20 (E178A)	+	+	0	II	Full length sequenced
F183A	++	+	0	II	
L184A	++++	+++	27	III	
I185A	++++	+++	23	III	Full length sequenced
L186A	++++	++++	66	I	
F193A	++++	++++	79	I	
F194A	++++	++++	67	I	
N204A	++++	++++	74	I	
G212A	++++	++++	73	I	
Y213A	++++	++++	75	I	
Y214A	++++	++++	69	I	
I218A	+++	+	43	III	
Y220A	+++	+	32	III	
F225A	+++	+	58	III	
M24 (E229A)	++++	++++	61	I	
Y232A	++++	++++	82	I	
L233A	++++	++++	82	I	
F234A	++++	++++	98	I	
Y241A	++++	++++	81	I	
L244A	++++	++++	72	I	
M26 (E245A)	++++	++++	55	I	
F248A	++++	++++	105	I	
Q251A	++++	++++	83	I	
F252A	++++	++++	74	I	
L253A	++++	++++	85	I	
Q255A	++++	++++	101	I	
L256A	++++	++++	99	I	
N257A	++++	++++	83	I	
M27 (E258A)	++++	++++	101	I	
I260A	++++	++++	110	I	
Y261A	++++	++++	111	I	
N268A	++++	++++	91	I	
L273A	++++	++++	85	I	
W275A	++++	++++	109	I	

<sup>a</sup> Class I, no major defect in protein expression or processing, virion incorporation, or viral entry (wild-type phenotype); class II, defect in protein expression or processing; class III, defect in virion incorporation; class IV, defect in viral entry.

the luciferase activities of the infected cells were determined as a measure of the GP-mediated viral entry. The  $\Delta$  mucin GP-mediated viral entry was two- to threefold more efficient than that of the wt GP (Fig. 1C), consistent with the previous reports (7).

Thirteen additional deletion mutants of GP that encompass the

N-terminal region of GP1 (residues 44 to 312), with the amino acid deletion in each mutant ranging from 18 to 26 residues, were generated (Fig. 1A). The construction of deletion mutants was guided by secondary-structure prediction programs (11) to minimize potentially deleterious effect of random deletion. However, among mutants  $\Delta$ 1 to  $\Delta$ 10, only the pre-GP forms were detect-

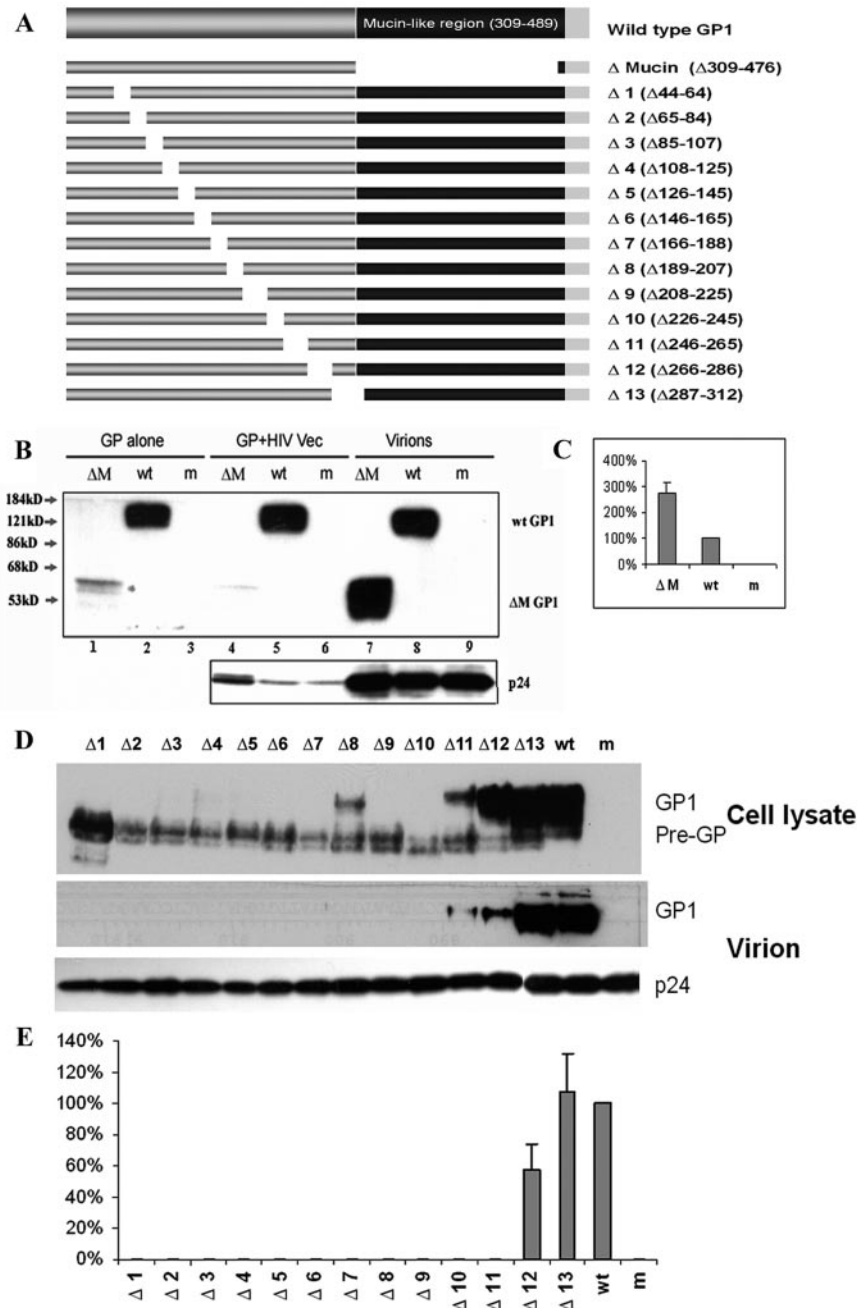


FIG. 1. Deletion analysis of GP1 in Ebola virus entry. (A) Diagrams of wild-type GP and deletion mutants  $\Delta$ mucin and  $\Delta 1$  to  $\Delta 13$ . The residues are numbered with the signal peptide (residues 1 to 32) according to GP0 of Ebola virus Zaire. (B) Expression and virion incorporation of  $\Delta$ mucin protein. GP alone, the  $\Delta$ mucin or wt GP gene was used in transfection. GP+HIV Vec, the GP gene and HIV vector were used in transfection. Lysates derived from transfected 293T cells were subjected to SDS-PAGE and Western blotting to detect wt or  $\Delta$ mucin GP1 proteins. Virions,  $\Delta$ mucin or wt GP proteins associated with HIV particles were detected by Western blotting. HIV p24 protein was detected in cell lysates or in the HIV particles. m, mock-transfected cells. (C) The infectivity of  $\Delta$ mucin GP-pseudotyped virus was measured by luciferase assay and is expressed as a percentage of wt infectivity (100%). Error bars indicate standard deviations. (D) Expression and virion incorporation of deletion mutants  $\Delta 1$  to  $\Delta 13$ . (E) Infectivity of  $\Delta 1$  to  $\Delta 13$  GP-pseudotyped virus by luciferase assay.

able, with exception of  $\Delta 8$  (Fig. 1D, cell lysate). Further, none of these deletion GPs was detected on the HIV virions (Fig. 1D, virion). These results suggest that these deletions adversely affected protein processing or folding, thus completely eliminating the ability of GP to mediate viral entry (Fig. 1E). Mutants  $\Delta 11$ ,  $\Delta 12$ , and  $\Delta 13$  displayed various degrees of protein processing and

viral incorporation. Only  $\Delta 12$  (60% of wt) and  $\Delta 13$  (110% of wt) were able to mediate relatively efficient viral entry (Fig. 1E), although  $\Delta 12$  appeared to be less efficient in incorporation. Together, the data from the deletion analysis have roughly defined that the N-terminal 230 residues (after the cleavage of the signal peptide) determine the viral entry function of GP1.

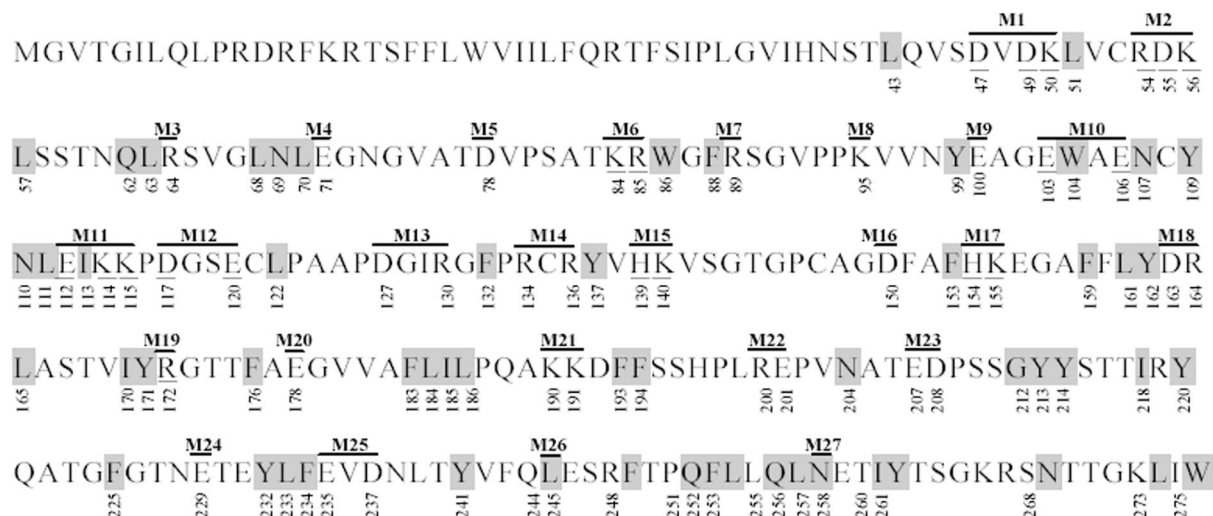


FIG. 2. Individual residues of GP1 targeted by alanine-scanning mutagenesis. The charged residues were replaced either individually (underlined) or in groups (labeled on top). The hydrophobic residues were targeted individually (shaded).

**Identification of residues of GP1 involved in protein folding and viral entry.** The deletion analysis of GP1 described above demonstrated that the N-terminal 230 residues are critical for GP1 function in viral entry. To identify the residues that are important in receptor recognition or protein folding, 103 individual residues within the N-terminal 230-residue region of GP1, which is conserved among different Ebola virus species, were targeted by alanine-scanning mutagenesis; these included 45 charged and 58 hydrophobic residues (Fig. 2). The choice of these residues for mutagenesis was based on the following considerations. First, sequence alignment of different Ebola virus species indicates that the N-terminal region of GP1 is highly conserved, suggesting that they adopt highly similar structures. Second, many of the conserved charged and hydrophobic residues are likely to be involved in protein folding or protein-protein interaction. Analysis of these mutants in protein expression or processing, incorporation onto HIV virions, and viral entry allowed us to roughly classify them into four groups, as follows: class I, no major defect in protein expression or processing, virion incorporation, or viral entry (wild-type phenotype); class II, defect in protein expression or processing; class III, defect in virion incorporation; and class IV, defect in viral entry. Thus, both class II and III mutants are referred to as being defective in GP folding and/or structure, while class IV mutants displayed no major defect in GP folding and/or structure but mediated viral entry at less than 20% of the wt GP level (Table 1).

**(i) Charged residues.** Twenty-seven mutants (M1 to M27) were generated in which either single or multiple charged residues of GP1 were replaced (Fig. 2). These mutations exerted different effects on protein processing or expression, virion incorporation, and viral entry (Fig. 3). Five mutants (M6, M10, M15, M17, and M20) were defective in protein expression or processing. We could detect little or no processed GP (GP1), although pre-GP forms could be easily detected (Fig. 3A [lane 6], C [lanes 1, 6, and 8], and E [lane 2]). As expected, none of these mutants was able to mediate viral

entry (Fig. 3B, D, and F). Four mutants (M2, M4, M12, and M14) were defective in GP incorporation into HIV virions (Fig. 3A [lanes 2 and 4] and C [lanes 3 and 5]). All of these mutants except M14 displayed complete elimination of viral entry function (Fig. 3B and D). Mutant M14 had a decreased level of incorporation but could mediate efficient viral entry (60% of the wt level). Another group of mutants (M1, M3, M8, and M11) did not display detectable defects in protein expression or virion incorporation but showed impaired viral entry function (<20% of wt level) (Fig. 3B and D). The remaining 14 mutants behaved like the wild type in GP expression, processing, incorporation into virions, and mediating viral entry (Fig. 3). It is interesting that none of the substitutions of the charged residues after residue 178 (M21 to M27) had a major adverse effect on the viral entry function of GP.

To identify the charged residues responsible for the defective phenotype, eight multiple substitution mutants (M1, M2, M6, M10, M11, M12, M15, and M17) were further characterized by single-amino-acid substitution (Fig. 4). It is clear that the processing defects of M6, M10, M15, and M17 were caused by the alteration of R85, E103, H139, and H154, respectively (Fig. 4A [lanes 8 and 9] and C [lanes 6 and 8]), because replacements of these residues alone displayed the same defective phenotype, while other individual replacements did not have a detectable adverse effect on protein processing (Fig. 4A [lanes 7 and 10] and C [lanes 7 and 9]). The incorporation defect in M12 was caused mainly by alteration of D117 (Fig. 4D, lanes 4, and 5). In contrast, the incorporation defect of M2 is likely caused by a combination of residues (R54, D55, and K56), since individual substitutions did not have any detectable effect on incorporation (Fig. 4A, lanes 4, 5, and 6). Similarly the viral entry defects of mutants M1 and M11 are likely due to multiple substitutions (Fig. 4A [lanes 1, 2, and 3] and C [lanes 1, 2, and 3]).

**(ii) Hydrophobic residues.** Among the 58 substitution mutants, 10 of them (L68, W86, Y99, Y109, F159, L161, Y162, Y171, F176, and F183) were defective in GP processing (Fig. 5). Most of these mutants, similar to the charged-residue pro-

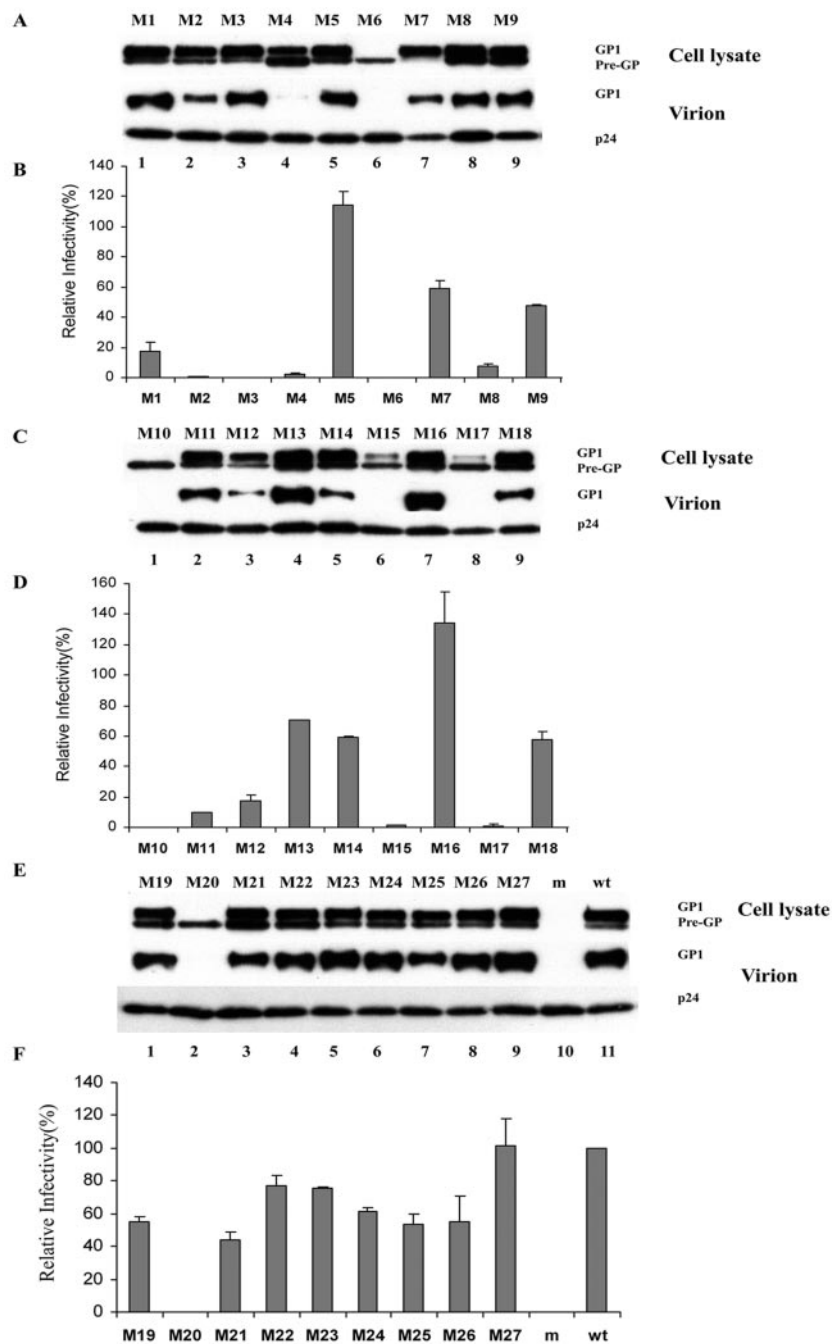


FIG. 3. Analysis of the charged-residue mutants M1 to M27. (A, C, and E) Protein expression and virion incorporation. (B, D, and F) Infectivity of the pseudotyped viruses. Error bars indicate standard deviations.

cessing-defective mutants described above, were unable to mediate viral entry, since the processed GP1 was barely or not detectable in cell lysates. GP incorporation onto HIV particles of 13 mutants was impaired to different degrees (Fig. 5). Two of the mutants (L111 and L122) showed defective incorporation, and thus they could not mediate viral entry. Six mutants (L43, L51, I113, F153, L184, and I185) had reduced level of incorporation, which explained why these mutants displayed reduced levels of infectivity. Sur-

prisingly, although the GP incorporation by five mutants (F132, Y137, I218, Y220, and F225) was severely reduced, they could still mediate relatively high level of viral entry (>25% of the wt level). Four mutants (L57, L63, F88, and I170) were identified as potentially entry defective because they were processed properly and incorporated onto virions efficiently (Fig. 5A [lanes 3, 5, and 10] and C [lane 6]), but they displayed reduced levels of viral entry (Fig. 5B and D). The remaining 31 mutants did not display major defects in

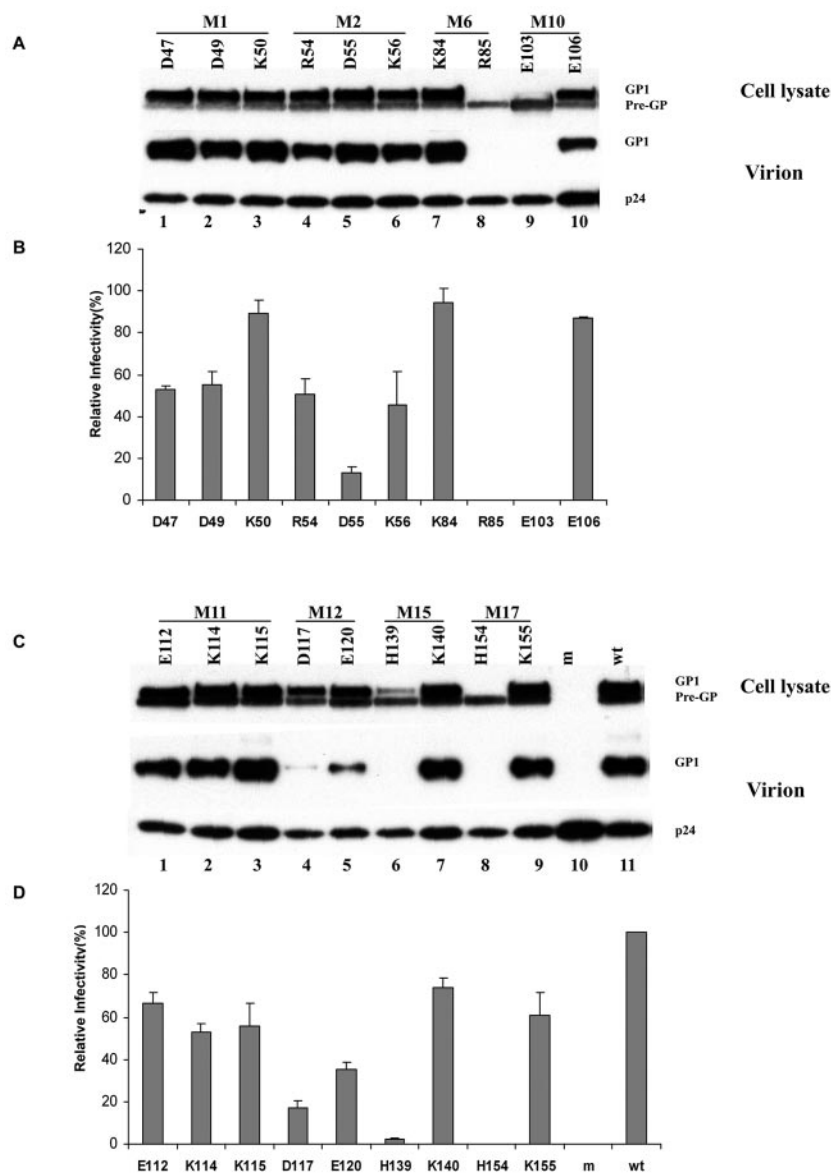


FIG. 4. Analysis of the charged-residue mutants D47 to K155. (A and C) Protein expression and virion incorporation. (B and D) Infectivity of the pseudotyped viruses. Error bars indicate standard deviations.

the characteristics and function of GP. In agreement with the results from substitutions of charged residues, the majority of substitutions after residue 185 did not have major effects on the viral entry function of GP.

**Characterization of putative receptor recognition residues.** Alanine substitution of each of the residues D55, L57, L63, R64, F88, K95, and I170 did not drastically affect GP expression, processing, or viral incorporation, suggesting that these substitutions did not cause a dramatic alteration in the native conformation of GP. Nevertheless, these mutants were shown to be defective in mediating efficient viral infection, giving less than 20% of wt GP activity (Fig. 3, 4, and 5). These residues are classified as putative receptor-binding residues which are likely involved in receptor binding and/or postbinding steps in Ebola virus entry. To better understand their roles in Ebola

virus entry, each of these residues was targeted for further mutational and functional analysis. The results are summarized below:

(i) **D55.** Substitution with another negatively charged residue (D55E) or with a positively charged residue (D55K) had no major effect on viral entry, while mutant D55A was impaired in viral infection (Fig. 6A and B). These results suggest that D55 may be surface exposed. However, since replacement of D55 by a positively charged residue did not exert an adverse effect on viral entry, it is unlikely that D55 is directly involved in receptor recognition.

(ii) **L57.** Substitution with other residues, including isoleucine (L57I, L57F, and L57K), inhibited the ability of GP to mediate viral infection (Fig. 6A and B), indicating a strict requirement for leucine at position 57.



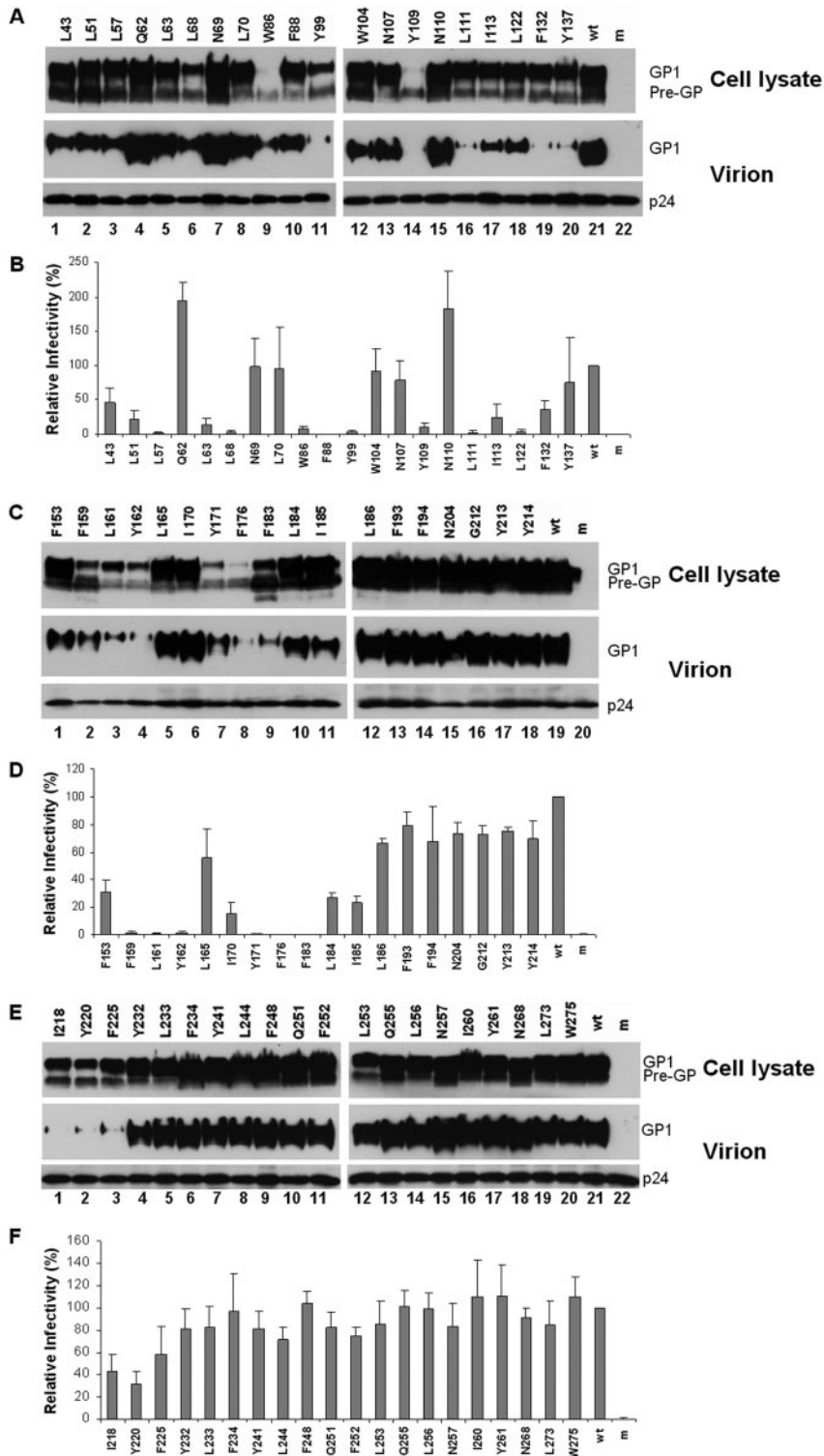


FIG. 5. Analysis of the hydrophobic-residue mutants. (A, C, and E) Protein expression and virion incorporation. (B, D, and F) Infectivity of the pseudotyped viruses. Error bars indicate standard deviations.

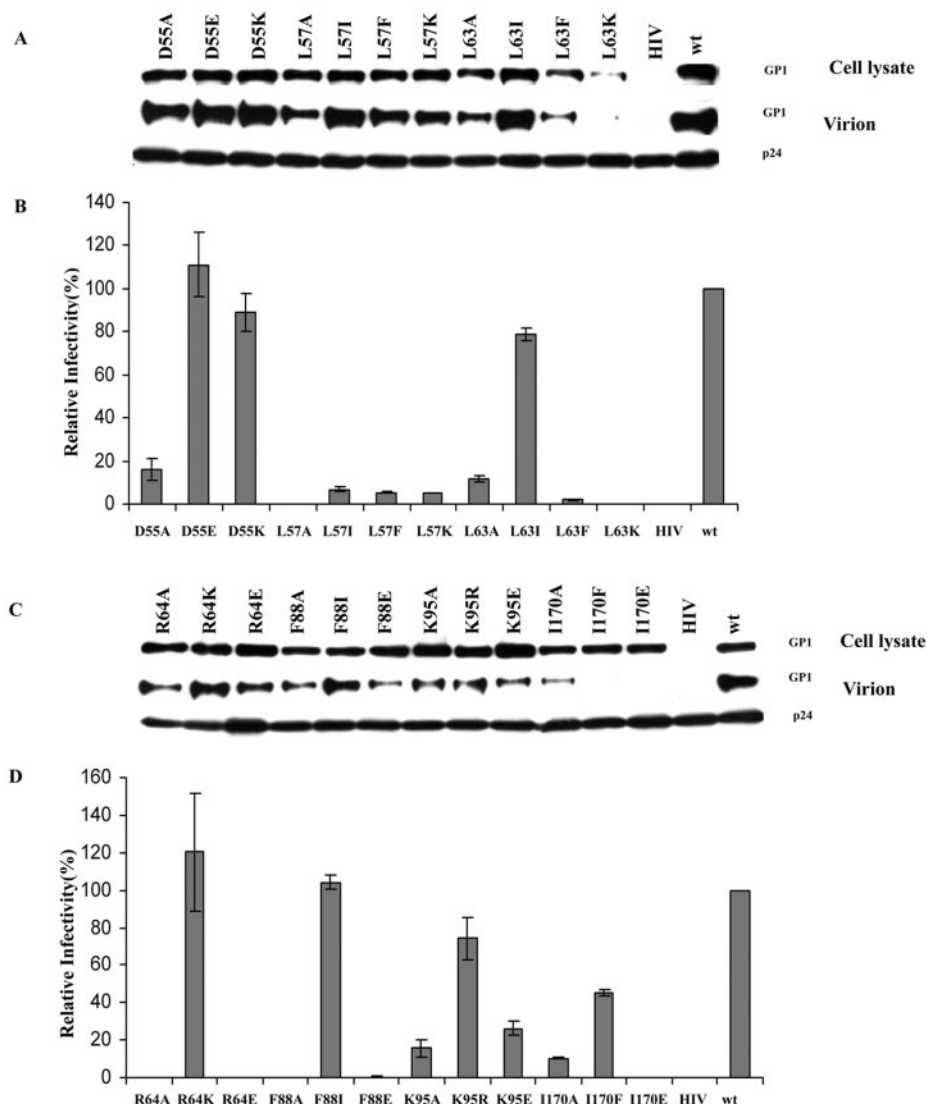


FIG. 6. Analysis of the putative receptor recognition residues. (A and C) Protein expression and viral incorporation. (B and D) Infectivity of the pseudotyped viruses. Error bars indicate standard deviations.

(iii) **L63**. Substitution with isoleucine (L63I) was well tolerated, while other substitutions (L63F, L63K, and L63A) impaired viral entry, likely due to structural alterations (Fig. 6A and B), suggesting a strict requirement for size and hydrophobicity in at position 63.

(iv) **R64**. Substitution with a positively charged residue (R64K) did not adversely affect GP function, while substitution with a negatively charged residue (R64E) inhibited GP function (Fig. 6C and D), suggesting that maintaining the positive charge at position 64 is critical.

(v) **F88**. Substitution with isoleucine (F88I) did not impair GP function, while substitution with a charged residue (F88E) rendered GP nonfunctional (Fig. 6C and D), suggesting a requirement for a hydrophobic residue at position 88.

(vi) **K95**. Substitution with another positively charged residue (K95R) had no major effect, while substitution with an oppositely charged residue (K95E) impaired GP function (Fig.

6C and D), suggesting a requirement for a positively charged residue at position 95.

(vii) **I170**. Even though substitution with an aromatic residue (I170F) at position 170 appeared to disrupt the proper native conformation of GP protein, as manifested by a low efficiency of viral incorporation (Fig. 6C), this mutant could still efficiently mediate viral entry (50% of the level of wt GP). However, substitution with a charged residue (I170E) eliminated GP structure integrity function (Fig. 6C and D). These data indicate a requirement for a hydrophobic residue at this position.

Together, these results are consistent with the notion that six residues (L57, L63, R64, F88, K95, and I170) are likely involved in receptor recognition and/or postbinding events.

## DISCUSSION

In this study, we performed a comprehensive analysis of the protein folding and viral entry function of Ebola virus GP1. A

series of deletion mutants encompassing the entire GP1 subunit and alanine substitutions of 103 individual residues in GP1 were analyzed for GP expression, processing, viral incorporation, and viral entry. The effects of each of the alanine substitutions on GP expression, virion incorporation, and viral entry function are summarized in Table 1. The results presented in this report suggest the N-terminal 150 residues in mature GP1 as the receptor-binding domain (RBD). Further, by alanine scanning of this region we have identified residues involved in protein folding and structure integrity. Of particular interest, we have identified several key residues which are likely involved in receptor recognition. These findings have important implications for understanding the structure and function of GP1 and providing potential therapeutic targets against Ebola virus infection.

Structural and functional information on the Ebola virus GP1 glycoprotein are important for understanding the entry mechanism of filoviruses and for development of therapeutic targets. We have identified 32 individual residues, including 23 hydrophobic and 9 charged residues, which are involved in GP1 folding or structure stability (Fig. 7A). These residues can be roughly classified into two groups: those whose replacement disrupted GP expression and processing and those whose replacement inhibited efficient incorporation of GP into viral particles. Although these two groups of mutants may represent protein folding defects at different stages of posttranslational modifications, they reflect overall perturbation of GP structure. Therefore, these residues must play critical roles in protein folding or structure. The information obtained in the present study provides a foundation for structural and functional studies of the role of the Ebola virus GP protein in viral entry in at least two aspects. First, the Ebola virus GP1 structure has not been reported to date, probably due to the technical difficulty of determining the structure of the glycoprotein. Thus, the structural information inferred from the present study on GP1 can be used as a guide for further biochemical and structural analyses of the glycoprotein. For example, instead of expressing and characterizing the full-length GP1 protein (approximately 500 residues in length), one could focus on the RBD region defined by the present study (about 150 residues), which is considerably smaller than the full-length GP1 protein and thus may be more amenable to biochemical and structural analyses, as long as the recombinant RBD maintains the biological activity of GP1 in receptor binding. Second, the information presented here will complement the structural analysis of the Ebola virus GP1 by either X-ray or nuclear magnetic resonance studies in the near future. This complementary nature of the functional and structural analyses of viral glycoproteins has been well demonstrated by studies of other viral glycoproteins, such as HIV gp120 and influenza virus HA1 (8, 20).

The GP1 subunit of Ebola virus glycoprotein is likely responsible for receptor binding. However, the receptor interaction region has not been identified. The present study demonstrates that two-thirds of GP1 (C terminus) is not essential for viral entry function. Our deletion analysis of GP1 revealed that ~40 residues N terminal to the mucin-like-region are also dispensable, thus assigning the viral entry function of GP1 to the N-terminal ~230 residues. Furthermore, mutational analysis of individual residues in this region suggested that residues

186 to 275 do not play a vital role in either protein folding or viral entry function of GP1, since none of the 38 mutations in this region had a major effect on the viral entry function and most of them (35 of 38) had no major effect on protein folding (Fig. 7A). In contrast, approximately half of the targeted residues in the N-terminal region (residues 33 to 185) are involved in protein folding and viral entry. These results strongly suggest that the N-terminal 150 residues (residues 33 to 185) of GP1, referred to as the RBD, determine receptor recognition. These results will facilitate characterization of GP-receptor interactions to elucidate the entry mechanism and host tropism of Ebola virus infection.

Substitution of six residues in the RBD did not affect GP expression, processing, or virion incorporation, suggesting that these residues are not involved in GP1 structure. Nevertheless, these mutants displayed an impaired phenotype in mediating viral entry (<20% of wt activity). Our results suggest that these residues are most likely involved in receptor recognition, and they will be referred to as putative receptor-binding residues. Interestingly three of them are localized in a 10-amino-acid stretch (Fig. 7A, L57, L63, and R64) potentially forming a receptor-binding pocket. Consistent with the notion that this region is involved in receptor recognition, we found that multiple substitutions of the charged residues (M1) near the putative binding pocket impaired viral entry (M1, D47A/D49A/K50A) (Fig. 3A and B), although they did not affect protein processing or virion incorporation. This putative receptor-binding region may be a potential therapeutic target against Ebola virus infection.

Sequence analysis of Ebola and Marburg virus GP1 subunits reveals that the overall sequence conservation is ~35% (identical plus conserved residues). It has been previously suggested that Ebola and Marburg viruses utilize different cellular receptors in viral entry (3). Comparison of the Ebola virus RBD and the corresponding region of Marburg virus GP1 indicate that the amino acid conservation is ~55%, approximately 20% higher than the overall conservation of GP1. Strikingly, 21 of 32 (66%) of the residues which were identified in this report to be involved in protein folding or structure of Ebola virus GP1 are conserved in Marburg virus (Fig. 7B), suggesting a highly conserved protein conformation in the receptor-binding domains of both Ebola and Marburg viruses. Furthermore, 6 of 6 (100%) putative receptor recognition residues identified in the Ebola virus RBD are also conserved in Marburg virus. Consistent with this notion, the secondary structure prediction analysis suggested that the Ebola virus RBD shares the same secondary structure with the corresponding region of Marburg virus GP1 (data not shown). These results strongly imply that the Marburg virus receptor-binding domain adapts a structure highly similar to that of the Ebola virus RBD and that the same or similar receptors are used by both Ebola and Marburg viruses during viral entry. Thus, it may be possible to develop the same or similar therapeutic inhibitors to block both Ebola and Marburg virus infection.

Expression of Ebola virus GP and sGP by using DNA vaccine or adenovirus vectors in animals elicits effective humoral and cell-mediated immune responses (22). Recently it has been demonstrated that cynomolgus macaques immunized with an adenovirus vector (encoding GP) were protected from Ebola virus infection, indicating that GP can be an effective

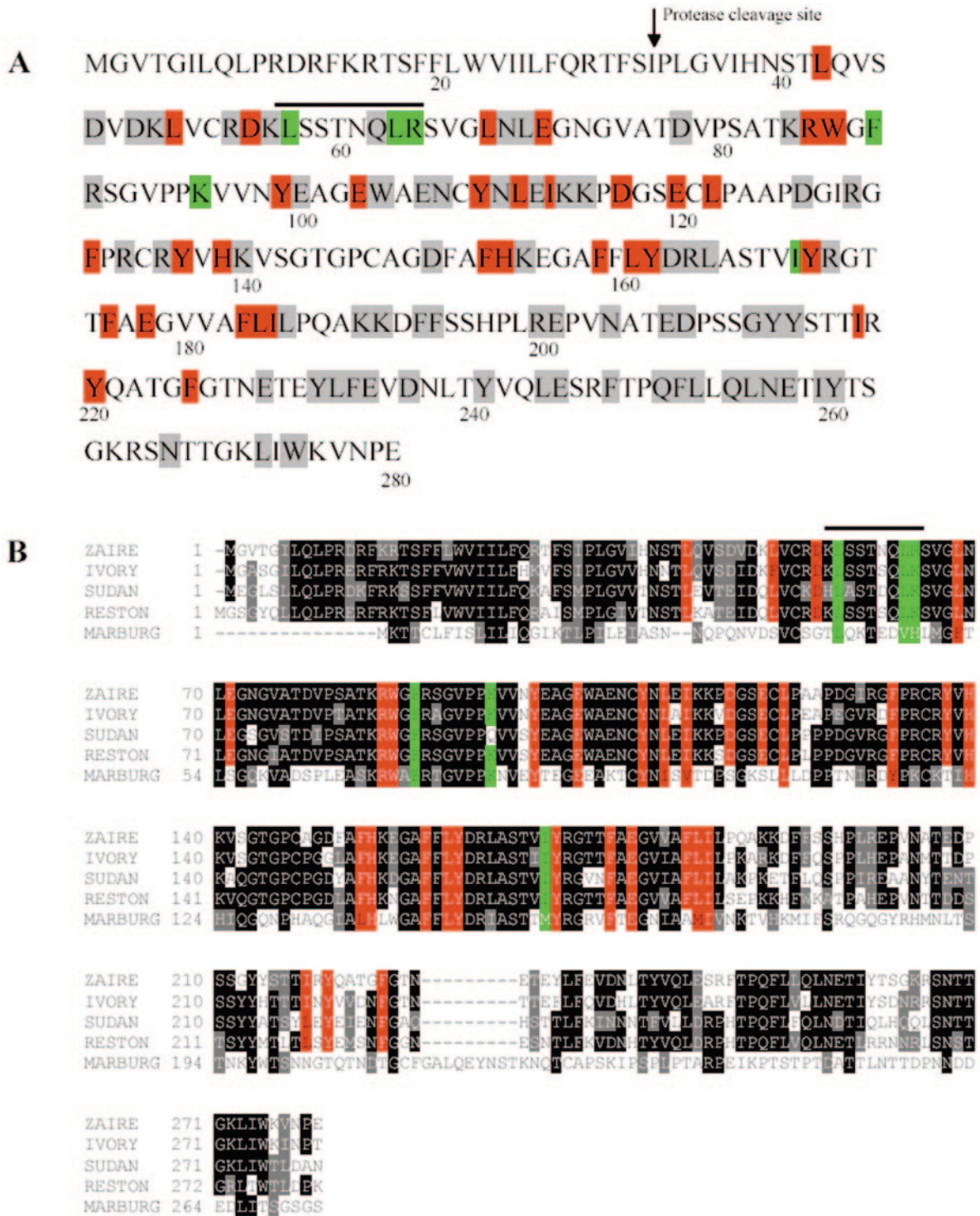


FIG. 7. Roles of individual residues of GP1 in protein structure and function of filoviruses. (A) Summary of functional analysis of individual residues of Ebola virus GP1. Red, residues involved in protein structure; green, putative receptor-binding residues; grey, no detectable defect in protein structure and function. The putative receptor-binding pocket is highlighted by a line on the top. (B) Sequence alignment of the receptor-binding domains of Ebola virus GP1 and Marburg virus GP1. Residues in red and green are implicated in protein folding and in receptor binding, respectively.

vaccine against Ebola virus infection (21). Furthermore, mice vaccinated with the Ebola virus-like particles generated by coexpression of GP and VP40 in mammalian cells were highly protected from lethal Ebola virus infection (28). These results suggest that GP alone or in combination with other Ebola virus proteins can be used as a prophylactic vaccine. However, because of the rapid course of disease progression (approximately 10 to 14 days after infection), the time (at least 1 month) required to get people immunized with the accelerated vaccine (if it proves to be effective in humans), and the unpredictable nature of outbreaks, other therapeutic approaches are urgently needed to combat Ebola virus infection. One attractive approach is to develop and design entry inhibitors to block Ebola virus infection and pathogenesis. Therefore, identification and characterization of peptides or small molecules which can bind to the receptor-binding pocket of Ebola virus GP1 and block GP1-receptor interactions may help facilitate the development of therapeutic treatment against Ebola virus infection in the near future.

In summary, the present study has mapped the putative receptor-binding domain to the N-terminal 150 amino acids of GP1 (residues 33 to 185). Many key residues involved in GP1 protein folding or structure and a putative receptor-binding pocket have been identified. These results provide important insights into GP1 folding, receptor usage, and potential therapeutic targets against filoviral infection.

#### ACKNOWLEDGMENTS

We thank Mary Kay Hart, U.S. Army Medical Research Institute of Infectious Diseases, for providing the monoclonal anti-GP antibodies and Michael Caffrey, University of Illinois at Chicago, for useful discussions. pNL4-3-Luc-R<sup>-</sup>E<sup>-</sup> was obtained from Nathaniel Landau through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH.

The laboratory research was supported by National Institutes of Health grants CA 092459 and AI 48056. L.R. was a recipient of the Schweppe Foundation Career Development Award.

#### REFERENCES

- Alvarez, C. P., F. Lasala, J. Carrillo, O. Muniz, A. L. Corbi, and R. Delgado. 2002. C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in *cis* and in *trans*. *J. Virol.* **76**:6841–6844.
- Chan, S. Y., C. J. Empig, F. J. Welte, R. F. Speck, A. Schmaljohn, J. F. Kreisberg, and M. A. Goldsmith. 2001. Folate receptor- $\alpha$  is a cofactor for cellular entry by Marburg and Ebola viruses. *Cell* **106**:117–126.
- Chan, S. Y., R. F. Speck, M. C. Ma, and M. A. Goldsmith. 2000. Distinct mechanisms of entry by envelope glycoproteins of Marburg and Ebola (Zaire) viruses. *J. Virol.* **74**:4933–4937.
- Connor, R. I., B. K. Chen, S. Choe, and N. R. Landau. 1995. Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. *Virology* **206**:935–944.
- Feldmann, H., V. E. Volchkov, V. A. Volchkova, U. Stroher, and H. D. Klenk. 2001. Biosynthesis and role of filoviral glycoproteins. *J. Gen. Virol.* **82**:2839–2848.
- He, J., S. Choe, R. Walker, P. Di Marzio, D. O. Morgan, and N. R. Landau. 1995. Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G<sub>2</sub> phase of the cell cycle by inhibiting p34cdc2 activity. *J. Virol.* **69**:6705–6711.
- Jeffers, S. A., D. A. Sanders, and A. Sanchez. 2002. Covalent modifications of the Ebola virus glycoprotein. *J. Virol.* **76**:12463–12472.
- Kwong, P. D., R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski, and W. A. Hendrickson. 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**:648–659.
- Lin, G., G. Simmons, S. Pohlmann, F. Baribaud, H. Ni, G. J. Leslie, B. S. Haggarty, P. Bates, D. Weissman, J. A. Hoxie, and R. W. Doms. 2003. Differential N-linked glycosylation of human immunodeficiency virus and Ebola virus envelope glycoproteins modulates interactions with DC-SIGN and DC-SIGNR. *J. Virol.* **77**:1337–1346.
- Malashkevich, V. N., B. J. Schneider, M. L. McNally, M. A. Milhollen, J. X. Pang, and P. S. Kim. 1999. Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-Å resolution. *Proc. Natl. Acad. Sci. USA* **96**:2662–2667.
- McGuffin, L. J., K. Bryson, and D. T. Jones. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* **16**:404–405.
- Sanchez, A., A. S. Khan, S. R. Zaki, G. J. Nabel, T. G. Ksiazek, and C. G. Peters. 2001. Filoviridae: Marburg and Ebola viruses, p. 1279–1304. *In* D. M. Knipe and P. M. Howley (ed.), *Fields virology*, 4th ed., vol. 1. Lippincott Williams & Wilkins, Philadelphia, Pa.
- Sanchez, A., M. P. Kiley, B. P. Holloway, and D. D. Aupepin. 1993. Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus. *Virus Res.* **29**:215–240.
- Sanchez, A., S. G. Trappier, B. W. Mahy, C. J. Peters, and S. T. Nichol. 1996. The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proc. Natl. Acad. Sci. USA* **93**:3602–3607.
- Sanchez, A., Z. Y. Yang, L. Xu, G. J. Nabel, T. Crews, and C. J. Peters. 1998. Biochemical analysis of the secreted and virion glycoproteins of Ebola virus. *J. Virol.* **72**:6442–6447.
- Simmons, G., J. D. Reeves, C. C. Grogan, L. H. Vandenberghe, F. Baribaud, J. C. Whitbeck, E. Burke, M. J. Buchmeier, E. J. Soilleux, J. L. Riley, R. W. Doms, P. Bates, and S. Pohlmann. 2003. DC-SIGN and DC-SIGNR bind Ebola glycoproteins and enhance infection of macrophages and endothelial cells. *Virology* **305**:115–123.
- Simmons, G., A. J. Rennekamp, N. Chai, L. H. Vandenberghe, J. L. Riley, and P. Bates. 2003. Folate receptor  $\alpha$  and caveolae are not required for Ebola virus glycoprotein-mediated viral infection. *J. Virol.* **77**:13433–13438.
- Simmons, G., R. J. Wool-Lewis, F. Baribaud, R. C. Netter, and P. Bates. 2002. Ebola virus glycoproteins induce global surface protein down-modulation and loss of cell adherence. *J. Virol.* **76**:2518–2528.
- Simon, J. H., R. A. Fouchier, T. E. Southerling, C. B. Guerra, C. K. Grant, and M. H. Malim. 1997. The Vif and Gag proteins of human immunodeficiency virus type 1 colocalize in infected human T cells. *J. Virol.* **71**:5259–5267.
- Skehel, J. J., and D. C. Wiley. 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* **69**:531–569.
- Sullivan, N. J., T. W. Geisbert, J. B. Geisbert, L. Xu, Z. Y. Yang, M. Roederer, R. A. Koup, P. B. Jahrling, and G. J. Nabel. 2003. Accelerated vaccination for Ebola virus haemorrhagic fever in non-human primates. *Nature* **424**:681–684.
- Sullivan, N. J., A. Sanchez, P. E. Rollin, Z. Y. Yang, and G. J. Nabel. 2000. Development of a preventive vaccine for Ebola virus infection in primates. *Nature* **408**:605–609.
- Takada, A., K. Fujioka, M. Tsuiji, A. Morikawa, N. Higashi, H. Ebihara, D. Kobasa, H. Feldmann, T. Irimura, and Y. Kawaoka. 2004. Human macrophage C-type lectin specific for galactose and N-acetylgalactosamine promotes filovirus entry. *J. Virol.* **78**:2943–2947.
- Takada, A., C. Robison, H. Goto, A. Sanchez, K. G. Murti, M. A. Whitt, and Y. Kawaoka. 1997. A system for functional analysis of Ebola virus glycoprotein. *Proc. Natl. Acad. Sci. USA* **94**:14764–14769.
- Takada, A., S. Watanabe, H. Ito, K. Okazaki, H. Kida, and Y. Kawaoka. 2000. Downregulation of beta1 integrins by Ebola virus glycoprotein: implication for virus entry. *Virology* **278**:20–26.
- Volchkov, V. E., S. Becker, V. A. Volchkova, V. A. Ternovoj, A. N. Kotov, S. V. Netsov, and H. D. Klenk. 1995. GP mRNA of Ebola virus is edited by the Ebola virus polymerase and by T7 and vaccinia virus polymerases. *Virology* **214**:421–430.
- Volchkov, V. E., H. Feldmann, V. A. Volchkova, and H. D. Klenk. 1998. Processing of the Ebola virus glycoprotein by the proprotein convertase furin. *Proc. Natl. Acad. Sci. USA* **95**:5762–5767.
- Warfield, K. L., C. M. Bosio, B. C. Welcher, E. M. Deal, M. Mohamadzadeh, A. Schmaljohn, M. J. Aman, and S. Bavari. 2003. Ebola virus-like particles protect from lethal Ebola virus infection. *Proc. Natl. Acad. Sci. USA* **100**:15889–15894.
- Weissenhorn, W., L. J. Calder, S. A. Wharton, J. J. Skehel, and D. C. Wiley. 1998. The central structural feature of the membrane fusion protein subunit from the Ebola virus glycoprotein is a long triple-stranded coiled coil. *Proc. Natl. Acad. Sci. USA* **95**:6032–6036.
- Weissenhorn, W., A. Carfi, K. H. Lee, J. J. Skehel, and D. C. Wiley. 1998. Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. *Mol. Cell* **2**:605–616.
- Wilson, J. A., M. Hevey, R. Bakken, S. Guest, M. Bray, A. L. Schmaljohn, and M. K. Hart. 2000. Epitopes involved in antibody-mediated protection from Ebola virus. *Science* **287**:1664–1666.
- Wool-Lewis, R. J., and P. Bates. 1998. Characterization of Ebola virus entry by using pseudotyped viruses: identification of receptor-deficient cell lines. *J. Virol.* **72**:3155–3160.
- Yang, Z. Y., H. J. Duckers, N. J. Sullivan, A. Sanchez, E. G. Nabel, and G. J. Nabel. 2000. Identification of the Ebola virus glycoprotein as the main viral determinant of vascular cell cytotoxicity and injury. *Nat. Med.* **6**:886–889.