JOURNAL OF VIROLOGY, May 1991, p. 2682–2685
0022-538X/91/052682-04$02.00/0

Vol. 65, No. 5

# NOTES

## The N-Terminal 31 Amino Acids of Human Immunodeficiency Virus Type 1 Envelope Protein gp120 Contain a Potential gp41 Contact Site

MONA IVEY-HOYLE,[1]* ROBERT K. CLARK,[2] AND MARTIN ROSENBERG[1,2]

*Departments of Gene Expression Sciences[1] and Cell Sciences,[2] SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, Pennsylvania 19406*

Received 26 October 1990/Accepted 28 January 1991

**We have compared the expression of full-length gp160 envelope protein from human immunodeficiency virus type 1 with that of a deletion mutant lacking the N-terminal 31 amino acids of the mature protein (gp160Δ32). The gp160 and gp160Δ32 proteins are processed to yield gp41 and gp120 or gp120Δ32, respectively. In contrast to full-length gp120, gp120Δ32 failed to associate with gp41 at the cell surface, despite conformational integrity as judged by soluble CD4 binding. Thus, the N-terminal 31 amino acids of gp120, which contain hyperconserved sequences, are likely involved in forming a contact site for gp41.**

The envelope glycoproteins of human immunodeficiency viruses (HIVs) are derived from a precursor protein, gp160, which is proteolytically cleaved to form the external envelope glycoprotein gp120 and the transmembrane glycoprotein gp41 (1, 21). The gp41 protein serves as an anchor for gp120 at the surface of the virus or infected cell and also directs membrane fusion with target cells (9). The gp120 protein targets the virus to $CD4^+$ cells by binding to the CD4 receptor on host cell surfaces (14, 15). Thus, virus, infected cells, or cells expressing recombinant envelope proteins at their surface are able to bind to and fuse with $CD4^+$ cells (9, 13, 17, 19).

The gp120 protein remains associated with gp41 at the cell surface by noncovalent interactions, and substantial amounts of gp120 are also shed into the medium (8, 15). The regions of gp120 and gp41 responsible for their association have not been fully defined. Mutational analyses have indicated that sites within the first 30 amino acids of gp41 and one site near the middle of gp41 are involved in association (4, 9). In addition, several insertion mutations ranging from amino acid residues 36 to 279 of the mature gp120 molecule (residues 66 to 309 of the precursor) have been shown to disrupt association with gp41 (9). This distribution of mutations in gp120 suggested that multiple contacts between the amino terminus of gp120 and gp41 may be responsible for the association of the proteins.

Previously we used a construct encoding a gp160 molecule with a short truncation at the N terminus to study HIV type 1 (HIV-1) *env* expression in *Drosophila* Schneider 2 cells (6). In the construct used, pMtt160Δ32, the *env* signal sequence and N-terminal 31 amino acids of the mature HIV-1 BH10 envelope protein were replaced with the signal sequence of the human tissue plasminogen activator (tPA) gene (Fig. 1). Cleavage of the tPA signal sequence from this type of fusion leaves four amino acids from tPA (Gly-Ala-Arg-Ser) preceding amino acid residue 32 (Asp) of the mature envelope protein (5). Transcription of *env* in the pMtt160Δ32 construct is controlled by the inducible *Drosophila* metallothionein (Mtn) promoter (7). Upon induction of the Mtn promoter in stably transfected cells, gp160Δ32 can be detected in association with cells by Western immunoblot analysis of whole-cell lysates (Fig. 2A). The gp160Δ32 appears to be processed, as evidenced by the accumulation of a gp120-sized cleavage product (gp120Δ32) in the culture supernatant (Fig. 2A). However, no processed gp120Δ32 can be detected in association with cells, even though pMtt160Δ32 contains all of the sequences previously shown by mutational analyses to be involved in gp120-gp41 association (4, 9).

The failure of gp120Δ32 to remain cell associated is in marked contrast to the retention of gp120 on virus-infected cells or cells expressing full-length recombinant envelope protein (9, 10, 17, 23). It is possible that expression of *env* in *Drosophila* cells or our fusion of *env* sequences to the tPA signal sequence may have generated a gp120 molecule incapable of association with gp41. Alternatively, the deletion of the first 31 amino acids of mature gp120 may be responsible for the failure of gp120Δ32 to remain cell associated. To determine whether these deleted sequences encompass a previously unidentified region required for association with gp41, the sequence encoding the 31 N-terminal amino acids was restored at the beginning of the *env* gene in pMtt160Δ32 so that the tPA signal sequence was fused to the first amino acid of the mature envelope protein (Fig. 1). This construct, pMtt160, was transfected into *Drosophila* cells in the same manner as pMtt160Δ32 to generate a stable cell line. After induction of the Mtn promoter, whole-cell lysates and culture supernatants were analyzed by Western blot analysis for envelope protein production (Fig. 2B). As seen with gp160Δ32, the full length gp160 is expressed and found present in the cell pellets. However, in contrast to gp120Δ32, full-length gp120 can now also be found to be cell associated in addition to being shed into the medium. Apparently, adding back the N-terminal 31 amino acids leads to a cell-associated form of gp120. We also note that a gp41-sized protein can be detected in whole-cell lysates of both the pMtt160Δ32 and pMtt160 cell lines by Western blot analysis

* Corresponding author.

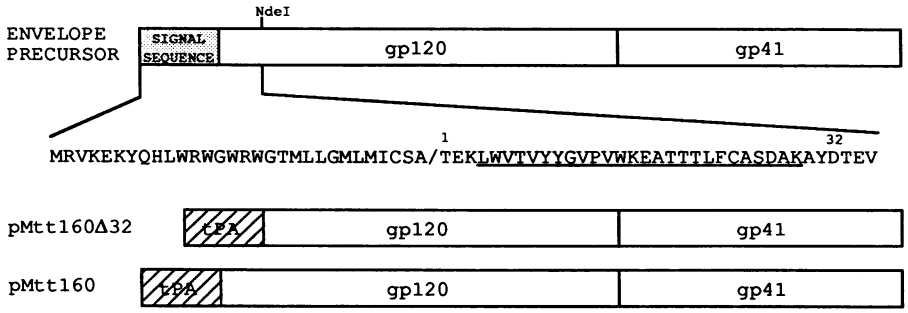


FIG. 1. Envelope expression plasmids. At the top, the HIV-1 envelope open reading frame is shown with the signal sequence shaded. The N-terminal sequence of the envelope precursor protein is represented underneath with the one-letter amino acid code. The underlined sequence is hyperconserved (15). The site of cleavage between the signal sequence and the mature N terminus of gp120 (1) is represented by a slash. The positions in the mature gp120 sequence numbered 1 and 32 represent the beginnings of the sequences that were fused in frame to the tPA signal sequence (striped box) to generate pMtt160 and pMtt160Δ32, respectively. The open reading frames shown for pMtt160Δ32 and pMtt160 are preceded by the *Drosophila* metallothionein promoter on a pBR322-based plasmid.

(Fig. 3). A summary of the cell-associated envelope proteins in the pMtt160Δ32 and pMtt160 cell lines detected by Western blot analysis is shown in Table 1.

The cellular location of envelope protein expressed in the induced pMtt160Δ32 and pMtt160 cell lines was investigated by indirect immunofluorescence assays using rabbit polyclonal antiserum specific for gp120 and then fluorescein-conjugated anti-rabbit immunoglobulin G antiserum (Cappel Laboratories, Malvern, Pa.). The results are summarized in Table 1. Both the pMtt160Δ32 and pMtt160 cell lines showed strong cytoplasmic staining, in contrast to nontransfected *Drosophila* S2 cells. However, membranous staining was detectable only on cells expressing full-length envelope protein from pMtt160. The fact that no surface fluorescence was detected with the pMtt160Δ32 cell line suggests that the cell-associated gp160Δ32 detected by Western blot analysis is not expressed at the cell surface. Thus, envelope expression in *Drosophila* cells appears identical to that in mammalian cells, where cleavable gp160 is not transported to the cell surface but rather is cleaved intracellularly (10, 18, 23). Assuming that the full-length gp160 produced by the pMtt160 cell line is also internal, the surface expression detected with this cell line likely results from gp120 being present at the surface. This interpretation is consistent with the fact that surface expression is detected only with the cell line (pMtt160) in which gp120 can be found to be cell associated by Western blot analysis. Thus, although both gp160Δ32 and full length gp160 appear to be processed appropriately, gp120Δ32 fails to remain cell associated while the full-length gp120 appears to stably associate with gp41 at the cell surface.

The association of full-length gp120 with gp41 at the cell surface rules out the possibility that some effect of expression in *Drosophila* cells (e.g., different glycosylation from that in mammalian cells) is responsible for the failure of gp120Δ32 to remain cell associated. We cannot rule out the formal possibility that gp120Δ32 fails to associate with gp41 because of an interfering effect of the four tPA-derived amino acids at its N terminus, even though these amino acids clearly do not interfere with gp41 association in the context of the full-length gp120 molecule. Thus, the four tPA-derived amino acids in and of themselves are compatible with gp41 association. The more likely explanation for the difference in gp41 association between gp120 and gp120Δ32 is the large deletion of 31 amino acids of envelope sequence.

The role of the N-terminal 31 amino acids of mature gp120

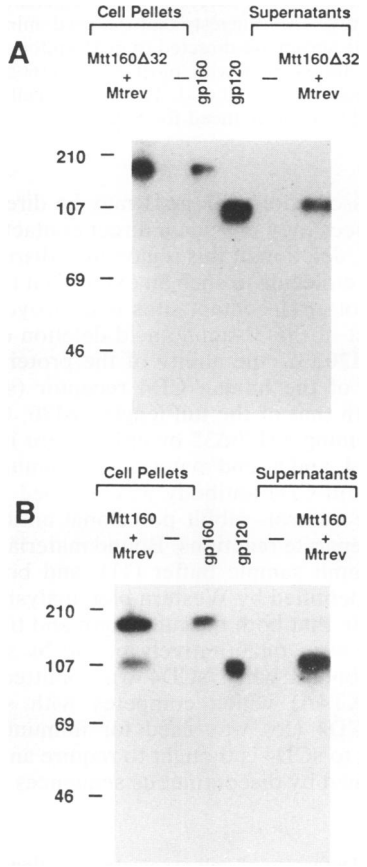


FIG. 2. Western blot analysis of envelope protein expression in stably transfected *Drosophila* cells. Cells cotransfected with the indicated envelope expression plasmid and a *rev* expression plasmid (pMtRev) or nontransfected cells (−) were induced for 5 days by the addition of 0.5 mM $CuSO_4$ to the culture medium. Cells were removed from culture by centrifugation, and whole-cell lysates from the cell pellets ($7 \times 10^5$ cells) or culture supernatant samples (15 μl) were analyzed by electrophoresis on 10% polyacrylamide gels followed by Western blot analysis using rabbit antiserum directed against gp120 and $^{125}$I-protein A (New England Nuclear). The center lanes in each panel contain gp160 and gp120 protein standards produced in recombinant baculovirus and *Drosophila* expression systems, respectively. Apparent molecular masses (in kilodaltons) are shown on the left. (A) Expression from the pMtt160Δ32 cell line. (B) Expression from the pMtt160 cell line.

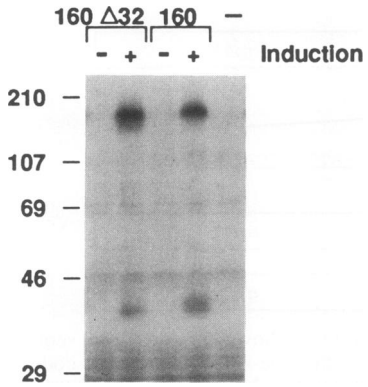


FIG. 3. Western blot analysis of gp160 and gp41 protein expression. Analysis of whole-cell lysates was carried out as described in the legend to Fig. 2, except that HIV-1-infected patient serum was used for detection. In previous blots with gp160 and gp120 standards, this patient serum detected gp160 efficiently but detected gp120 inefficiently. This suggests that the predominant *env*-specific antibodies in the serum are directed to gp41 epitopes. The cell lines analyzed were transfected with pMtt160Δ32 (160Δ32) or pMtt160 (160) or were not transfected (—). Transfected cells were analyzed both uninduced (−) and induced for 5 days (+).

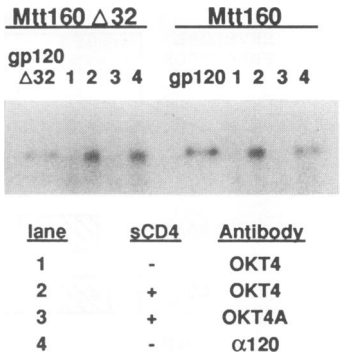


FIG. 4. Specific binding of secreted gp120Δ32 and gp120 to sCD4. Culture supernatants containing gp120Δ32 or gp120 secreted from the pMtt160Δ32 or pMtt160 cell line, respectively, were incubated in the presence (+) or absence (−) of sCD4 and then immunoprecipitated with antibody as indicated in the chart. Bound material was analyzed by Western blot analysis as described in the legend to Fig. 2. Lanes marked gp120Δ32 and gp120 contain untreated culture supernatants equal in volume to those used in the immunoprecipitation reactions.

in allowing association with gp41 may be direct or indirect. These sequences may provide a direct contact site for gp41. Alternatively, deletion of this region may disrupt the conformation of the molecule to such an extent that the appropriate arrangement of gp41 contact sites is destroyed. To investigate the effect of the 31-amino-acid deletion on the conformation of gp120Δ32, the ability of the protein to bind to a soluble form of the human CD4 receptor (sCD4 [3]) was compared with that of the full length gp120. Culture supernatants containing gp120Δ32 or gp120 were incubated with 0.5 μg of sCD4, and bound material was immunoprecipitated with excess anti-CD4 antibody as described elsewhere (5). As a positive control, rabbit polyclonal antibody to gp120 was used in separate reactions. Bound material was eluted at 90°C in Laemmli sample buffer (11), and bound envelope protein was identified by Western blot analysis (Fig. 4). The results indicate that both the full-length and truncated envelope proteins were quantitatively bound by sCD4. Neither protein was bound when sCD4 was omitted or when an antibody (OKT4A) which competes with specific gp120 binding to sCD4 (15) was used for immunoprecipitation. Since binding to sCD4 is thought to require an intact tertiary structure formed by discontinuous sequences in gp120 (2, 9) and since the sCD4 binding of gp120Δ32 is indistinguishable from that of full-length gp120, the 31-amino-acid deletion does not disrupt the overall conformation of the protein.

Other means by which the 31-amino-acid deletion may indirectly affect gp120-gp41 association include altering the extent or actual site of gp160 cleavage. We do not detect any significant differences in the mobility of gp41 produced from the full-length and deleted constructs (Fig. 3). However, there is a small difference in the amount of gp41 detected, but it seems unlikely that this difference could account for the complete absence of cell-associated gp120Δ32. It seems more likely that full-length gp120 remains cell associated, in contrast to gp120Δ32, because the N-terminal 31 amino acids of the protein are somehow involved in the formation of a critical contact site for gp41.

Alignment of the amino acid sequences at the N termini of 19 different HIV-1 isolates (16) reveals a hyperconserved stretch of 26 amino acids in which 21 of the positions are invariant and 4 are rarely substituted, with conservative changes only. Across this region (amino acids 4 to 29 of mature gp120), the sequence of the HIV-1 BH10 isolate used in this study (Fig. 1) is identical to the consensus. The role of this highly conserved region within the envelope has not been previously defined. Clearly, we and others have shown that it does not affect CD4 binding (2, 12, 20). Our finding that deletion of this hyperconserved region completely abolishes association of gp120 at the cell surface suggests that the role of this conserved sequence is to provide important information for stable association with gp41. Others have shown that a small insertion very near this region, at amino acid 36 of the mature envelope protein (amino acid 66 of the precursor), abolishes gp120-gp41 association in mammalian cells (9). Mutations in other regions within the first 300 amino acids of mature gp120 downstream of the N-terminal hyperconserved region also have been shown to disrupt gp120-gp41 association without affecting CD4 binding (9). Thus, the association between gp120 and gp41 appears to depend on multiple contacts.

We find that the pMtt160Δ32 and pMtt160 *Drosophila* cell lines fail to fuse with $CD4^+$ cells. Other insect cell lines (*Spodoptera frugiperda*) have been shown to fuse with $CD4^+$ cells upon expression and cleavage of gp160, suggest-

TABLE 1. Detection of cell-associated envelope proteins by immunoreactivity

| Plasmid transfected | Reactivity in: | | | | |
|---|---|---|---|---|---|
| | Immunoblot[a] | | | Immunofluorescence assay | |
| | gp160Δ32 or gp160 | gp41 | gp120Δ32 or gp120 | Cytoplasmic[b] | Membranous[c] |
| pMtt160Δ32 | + | + | − | + | − |
| pMtt160 | + | + | + | + | + |
| None | − | − | − | − | − |

[a] Data from Fig. 2 and 3.
[b] Immunoreactivity of anti-gp120 antibody with cytocentrifuge preparations fixed in absolute methanol.
[c] Immunoreactivity of anti-gp120 antibody with viable-cell suspensions.

ing that insect cells cleave gp160 into functional gp120 and gp41 molecules capable of directing fusion (22). It is possible that the *Drosophila* S2 cell line lacks some accessory factor needed for *env*-mediated fusion with $CD4^+$ cells. Alternatively, it may be that the gp160 cleavage efficiency, the density of gp120-gp41 at the cell surface, or the presentation of the molecules at the surface is inappropriate to allow fusion to occur.

Our data suggest that the N-terminal 31 amino acids of mature gp120 are critical for gp120-gp41 association and thus presumably for viral infectivity. The evidence that these sequences may provide a contact site for gp41 provides a plausible role for a hyperconserved region present at the extreme N terminus of the HIV-1 envelope protein.

We thank Elizabeth Lee for excellent technical assistance and James Hoxie (University of Pennsylvania) for the generous gift of HIV-1-infected patient serum.

This work was supported by grant AI-24845-02 from the National Institutes of Health.

## REFERENCES

1. **Allan, J. S., J. E. Coligan, F. Barin, M. F. McLane, J. G. Sodroski, C. A. Rosen, W. A. Haseltine, T. H. Lee, and M. Essex.** 1985. Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. Science **228**:1091–1094.
2. **Cordonnier, A., Y. Rivière, L. Montagnier, and M. Emerman.** 1989. Effects of mutations in hyperconserved regions of the extracellular glycoprotein of human immunodeficiency virus type 1 on receptor binding. J. Virol. **63**:4464–4468.
3. **Deen, K. C., J. S. McDougal, R. Inacker, G. Folena-Wasserman, J. Arthos, J. Rosenberg, P. J. Maddon, R. Axel, and R. W. Sweet.** 1988. A soluble form of CD4 (T4) protein inhibits AIDS virus infection. Nature (London) **331**:82–84.
4. **Freed, E. O., D. J. Myers, and R. Risser.** 1990. Characterization of the fusion domain of the human immunodeficiency virus type 1 envelope glycoprotein gp41. Proc. Natl. Acad. Sci. USA **87**:4650–4654.
5. **Ivey-Hoyle, M., J. S. Culp, M. A. Chaikin, B. D. Hellmig, T. J. Matthews, R. W. Sweet, and M. Rosenberg.** 1991. Envelope glycoproteins from biologically diverse isolates of immunodeficiency viruses have widely different affinities for CD4. Proc. Natl. Acad. Sci. USA **88**:512–516.
6. **Ivey-Hoyle, M., and M. Rosenberg.** 1990. Rev-dependent expression of human immunodeficiency virus type 1 gp160 in *Drosophila melanogaster* cells. Mol. Cell. Biol. **10**:6152–6159.
7. **Johansen, H., A. van der Straten, R. Sweet, E. Otto, G. Maroni, and M. Rosenberg.** 1989. Regulated expression at high copy number allows production of a growth-inhibitory oncogene product in *Drosophila* Schneider cells. Genes Dev. **3**:882–889.
8. **Kieny, M. P., G. Rautmann, D. Schmitt, K. Dott, S. Wain-Hobson, M. Alizon, M. Girard, S. Chamaret, A. Laurent, L. Montagnier, and J. P. Lecocq.** 1986. AIDS virus env protein expressed from a recombinant vaccinia virus. Bio/Technology **4**:790–795.
9. **Kowalski, M., J. Potz, L. Basiripour, T. Dorfman, W. C. Goh, E. Terwilliger, A. Dayton, C. Rosen, W. Haseltine, and J. Sodroski.** 1987. Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. Science **237**:1351–1355.
10. **Kozarsky, K., M. Penman, L. Basiripour, W. Haseltine, J. Sodroski, and M. Krieger.** 1989. Glycosylation and processing of the human immunodeficiency virus type 1 envelope protein. J. Acquired Immune Defic. Syndr. **2**:163–169.
11. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
12. **Lasky, L. A., G. Nakamura, D. H. Smith, C. Fennie, C. Shimasaki, E. Patzer, P. Berman, T. Gregory, and D. J. Capon.** 1987. Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with CD4 receptor. Cell **50**:975–985.
13. **Lifson, J. D., G. R. Reyes, M. S. McGrath, B. S. Stein, and E.G. Engleman.** 1986. AIDS retrovirus induced cytopathology: giant cell formation and involvement of CD4 antigen. Science **232**: 1123–1127.
14. **Maddon, P. J., A. G. Dalgleish, J. S. McDougal, P. R. Clapham, R. A. Weiss, and R. Axel.** 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. Cell **47**:333–348.
15. **McDougal, J. S., M. S. Kennedy, J. M. Sligh, S. P. Cort, A. Mawle, and J. K. A. Nicholson.** 1986. Binding of the HTLV III/LAV to $T4^+$ T cells by a complex of the 110K viral protein and the T4 molecule. Science **231**:382–385.
16. **Myers, G., A. B. Rabson, J. A. Berzofsky, T. F. Smith, and F. Wong-Staal (ed.).** 1990. Human retroviruses and AIDS. Los Alamos National Laboratory, Los Alamos, N.Mex.
17. **Sodroski, J., W. C. Goh, C. Rosen, K. Campbell, and W. A. Haseltine.** 1986. Role of the HTLV-III/LAV envelope in syncytium formation and cytopathicity. Nature (London) **322**:470–474.
18. **Stein, B. S., and E. G. Engleman.** 1990. Intracellular processing of the gp160 HIV-1 envelope precursor. J. Biol. Chem. **265**: 2640–2649.
19. **Stein, B. S., S. D. Gowda, J. D. Lifson, R. C. Penhallow, K. G. Bensch, and E. G. Engleman.** 1987. pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. Cell **49**:659–668.
20. **Syu, W.-J., J.-H. Huang, M. Essex, and T.-H. Lee.** 1990. The N-terminal region of the human immunodeficiency virus envelope glycoprotein gp120 contains potential binding sites for CD4. Proc. Natl. Acad. Sci. USA **87**:3695–3699.
21. **Veronese, F. D., A. L. DeVico, T. D. Copeland, S. Oroszlan, R. C. Gallo, and M. G. Sarngadharan.** 1985. Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. Science **229**:1402–1404.
22. **Wells, D. E., and R. W. Compans.** 1990. Expression and characterization of a functional human immunodeficiency virus envelope glycoprotein in insect cells. Virology **176**:575–586.
23. **Willey, R. L., J. S. Bonifacino, B. J. Potts, M. A. Martin, and R. D. Klausner.** 1988. Biosynthesis, cleavage, and degradation of the human immunodeficiency virus 1 envelope glycoprotein gp160. Proc. Natl. Acad. Sci. USA **85**:9580–9584.