JOURNAL OF VIROLOGY, Apr. 1992, p. 1856–1865
0022-538X/92/041856-10$02.00/0

Vol. 66, No. 4

Gag Proteins of the Highly Replicative MN Strain of Human Immunodeficiency Virus Type 1: Posttranslational Modifications, Proteolytic Processings, and Complete Amino Acid Sequences

LOUIS E. HENDERSON,[1]* MICHELLE A. BOWERS,[1] RAYMOND C. SOWDER II,[1] STEFAN A. SERABYN,[1] DONALD G. JOHNSON,[1] JULIAN W. BESS, JR.,[1] LARRY O. ARTHUR,[1] DUNCAN K. BRYANT,[2] AND CATHERINE FENSELAU[2]

AIDS Vaccine Program, Program Resources, Incorporated/DynCorp, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201,[1] and Department of Chemistry and Biochemistry, University of Maryland, Baltimore County, Baltimore, Maryland 21228[2]

Received 3 September 1991/Accepted 17 December 1991

The MN strain of human immunodeficiency virus type 1 was grown in H9 cells, concentrated by centrifugation, and disrupted, and proteins were purified by reversed-phase high-pressure liquid chromatography. Complete amino acid sequences were determined for the mature Gag proteins, showing natural proteolytic cleavage sites and the order of proteins (p17-p24-p2-p7-p1-p6) in the Gag precursors. At least two sequence variants of p24 and eight sequence variants of p17 were detected. The two most abundant variants of p24 and p17 represented at least 50% ± 5% and 20% ± 5% of their totals, respectively. These data suggest heterogeneity in the virus population, with 50% of the total virus containing the most abundant forms of p17 and p24 and 20% of the virus containing the second most abundant forms. The Gag precursors of these suggested viruses differ from each other by only 3 amino acid residues but differ from the precursors predicted by the published MN proviral DNA sequence by 10 residues. Electrospray ionization mass spectrometry analysis of the purified p24 forms showed that the measured molecular weight of the protein was 200 ± 50 atomic mass units greater than the calculated molecular weight. The source of additional mass for the p24 forms was not determined, but the observation is consistent with previous suggestions that the protein is phosphorylated. Greater than 98% of the total recovered p17 was myristylated at the N-terminal glycine residue, and the measured molecular weights (as determined by electrospray ionization mass spectrometry) of the most abundant forms were within 3 atomic mass units of the calculated molecular weights (15,266).

Human immunodeficiency virus type 1 (HIV-1) is the etiologic agent of AIDS (1, 25, 29), and the MN strain of HIV-1 (HIV-1_{MN}) (13) has been shown to be a prevalent prototype in the U.S. population, both by serological studies (7) and by virus isolation and sequencing of the V3 region of the *env* gene (23). Because of its prevalence, HIV-1_{MN} has been suggested as the prototype virus for the development of vaccines and other antiviral strategies. Standardized stocks of this strain will be needed as challenge reagents for testing the effectiveness of various developed procedures; it is highly desirable that these stocks be well characterized. Challenge stocks of HIV-1_{MN} grown in H9 cells have been prepared and are being evaluated at the biological level. At the molecular level, at least one complete infectious provirus has been isolated from cells infected with HIV-1_{MN}, and its nucleotide sequence has been determined (28). However, infected cells generally contain multiple sites of proviral integration, and it is difficult to determine which of the integrated proviruses produces the most abundant virus in tissue cultures. One method for characterizing the most abundant virus is to isolate viral proteins and determine their amino acid sequences.

The HIV-1 viral genome contains three large open reading frames, designated *gag*, *pol*, and *env*, that code for proteins ultimately destined for incorporation into the mature virus. Gag and Pol products together with RNA make up the viral core, and Env products are part of the lipid envelope. Pol proteins are enzymes necessary for viral replication. Gag proteins are often referred to as structural proteins and represent a major portion of the core structure. The proteins analyzed here are products of the *gag* gene purified directly from HIV-1_{MN} grown in tissue cultures. The primary translational product of the *gag* gene is a polyprotein precursor, designated $Pr55^{gag}$, that is ultimately cleaved by the viral protease to the mature Gag proteins found in the virus. Here we show that $Pr55^{gag}$ of HIV-1_{MN} is cleaved to six products, including the matrix antigen (MA), core antigen (CA), nucleocapsid (NC) protein, and three peptides, designated p1, p2, and p6, and the order of the products in the precursor is MA-CA-p2-NC-p1-p6, as previously suggested (15, 18). We report on the purification and complete chemical structural analysis of the six mature Gag proteins purified directly from HIV-1_{MN} viral particles.

MATERIALS AND METHODS

Virus preparation. H9 cells chronically infected with HIV-1_{MN} were provided by R. C. Gallo (Laboratory of Tumor Cell Biology, Division of Cancer Etiology, National Cancer Institute, Bethesda, Md.) and were grown in roller bottles under biosafety level 3 laboratory conditions at the National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, Md. Virus production was increased approximately 50-fold by supplementing infected H9 cells with an equal number of uninfected H9 cells after each harvest (2). Virus was harvested from the culture medium every 3 to 4 days, isolated by continuous-flow sucrose

* Corresponding author.

density centrifugation (34), and further concentrated by ultracentrifugation (100,000 × *g*). Concentrated virus was resuspended in 0.01 M Tris hydrochloride (pH 7.2)–0.1 M NaCl–1.0 mM EDTA to a final protein concentration of 8 to 20 mg/ml and stored frozen at −70°C.

As a preliminary step in protein purification, concentrated virus was disrupted and reduced under biosafety level 3 conditions as follows. Solid guanidine hydrochloride (Pierce) was added to suspensions of concentrated virus (1.53 g/ml of virus suspension) to yield a final concentration of 0.76 g/ml (saturated at room temperature), the pH was adjusted to 8.5 by the addition of concentrated Tris hydrochloride buffer, and 2-mercaptoethanol was added to a final concentration of 2% (vol/vol). The resulting clear solution was transferred to a clean sterile tube and removed from the biosafety level 3 environment.

Protein and peptide purification. All samples for reversed-phase high-pressure liquid chromatography (rp-HPLC) were dissolved in saturated guanidine hydrochloride containing 2% 2-mercaptoethanol at pH 8.5 and sonicated at 50°C for 15 min before injection into the chromatographic column. Separations were performed on μBondapak C18 (Waters Associates) rp-HPLC supports by use of a liquid chromatograph (Pharmacia LKB) equipped with a rapid spectral detector (model 2140). Elutions were accomplished with 0.05% (vol/vol) trifluoroacetic acid at pH 2 and with a gradient of increasing acetonitrile concentrations. Eluted proteins and peptides were detected by UV absorption at 206, 280, and 294 nm and collected, and the solvents were removed by lyophilization.

Endoproteinase digestion. Proteins to be digested were dissolved in 50 mM NH_4HCO_3 at pH 8 to a final concentration of 1 mg/ml. When a protein was initially insoluble, the pH was raised to 11 (concentrated NaOH) to solubilize the protein and then adjusted to 8 to 9 (concentrated HCl) before addition of the enzyme and buffer. Digestions were initiated by adding the endoproteinase to a final concentration of 0.02 mg/ml and were continued for 18 h at room temperature. When a precipitate developed during the course of the digestion, the process was repeated for a second 18-h period. Digestions were stopped by adding concentrated hydrochloric acid to a final pH of 2, and the peptides were separated by rp-HPLC on μBondapak C18 columns (3.9 by 300 mm). The endoproteinases used were endoproteinase Glu-C, endoproteinase Lys-C, and trypsin (Boehringer, Mannheim, Germany).

Amino acid analysis. Approximately 1 μg of purified protein or peptide was hydrolyzed in 200 μl of 6 M hydrochloric acid containing 0.1% phenol in vacuo at 110°C for 24 h. The samples were dried in vacuo and analyzed on a Beckman model 6300 automated amino acid analyzer in accordance with the manufacturer's recommendations.

Gel electrophoresis. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed as previously described (22).

Edman degradation. Automatic N-terminal Edman degradations were performed on approximately 1-nmol samples of purified protein or peptide by use of a pulsed, liquid-phase protein sequencer (model 477A; Applied Biosystems Inc.) equipped with an on-line phenylthiohydantoin analyzer (model 120A) in accordance with the manufacturer's recommendations.

FAB mass spectrometry. Fast atom bombardment (FAB) mass spectrometry of Gag proteins was performed by use of the first two sectors of an HX110/HX110 mass spectrometer (JEOL, Tokyo, Japan) at an accelerating voltage of 10 kV, with 100-Hz filtering, and at a resolution of 1:1,000 (10% valley). A JEOL FAB gun was operated at 6 kV with xenon as the FAB gas. Spectra were recorded with a JEOL DA5000 data system.

Tandem mass spectrometry. FAB tandem mass spectrometry was carried out by use of all four sectors on the JEOL HX110/HX110 instrument. Precursor ions were mass selected in the first mass spectrometer, and linked B/E scans were made in the second mass spectrometer. High-energy collision-induced dissociations occurred in a chamber between the two mass spectrometers (4), with helium used as the collision gas at a pressure sufficient to attenuate the precursor ion beam by 80%. The collision cell was floated at 4 kV. The collision-induced dissociation spectra were recorded at a resolution of 1:1,000 and with 100-Hz filtering. For both scanning modes, 50 to 100 pmol of peptide in 0.1% trifluoroacetic acid (1 μl) and 3-nitrobenzyl alcohol (1 μl) were used for the analysis.

The computer program RESIDUES, written by David Heller (Middle Atlantic Mass Spectrometry Center, Johns Hopkins University, Baltimore, Md.), was used to assist with the interpretation of tandem mass spectrometry data, while the computer program MacProMass (written by Terry D. Lee and Sunil Vemuri, Beckman Research Institute, City of Hope, Duarte, Calif.) was used to identify p17 proteolysis peptides in the conventional magnetic scans.

ESI-MS. Electrospray ionization mass spectrometry (ESI-MS) was performed with a Vestec electrospray source fitted to a Hewlett-Packard 5988A quadrupole mass spectrometer. Sample p17 (from pool A) was dissolved in an aqueous solution containing 1.5% acetic acid and 50% methanol. This solution was introduced into the source with a Harvard 22 syringe infusion pump.

RESULTS

Isolation of HIV-1_{MN} proteins. As part of an ongoing program to characterize HIV-1_{MN} and to prepare challenge stocks for future studies, the virus was produced in the AIDS Vaccine Program virus production facility at the National Cancer Institute-Frederick Cancer Research and Development Center by cocultivating infected and uninfected H9 cells. Compared with the amount of virus produced from H9 cells chronically infected with HIV-1_{MN}, this cocultivation method results in approximately a 50-fold increase in the amount of virus recovered and has been adopted as a standard production method (2). Concentrated purified virus was disrupted and solubilized in saturated guanidine hydrochloride (pH 8.5) containing 2-mercaptoethanol, and the resulting clear solution was injected into a high-pressure liquid chromatograph for separation and purification of the viral proteins. Figure 1 shows a typical chromatogram for the rp-HPLC separation of the proteins associated with the purified virus. Each peak of eluted protein was analyzed by SDS-PAGE, amino acid analysis, and N-terminal Edman degradation to qualitatively identify the purified proteins. Many of the peaks were found to contain cellular proteins, and these will be discussed in a separate communication (18a). The peaks containing purified HIV-1_{MN} Gag proteins p17, p24, p2, p7, p1, and p6 are indicated in Fig. 1. Protein in each of the labeled peaks (Fig. 1) was analyzed by SDS-PAGE, and the results for peaks p7, p1, p6, p17a through p17e, and p24 are shown in Fig. 2. Peak p17f was not included in this gel but yielded results indistinguishable from those yielded by p17e (data not shown); peak p2 yielded no detectable bands in the SDS-PAGE analysis

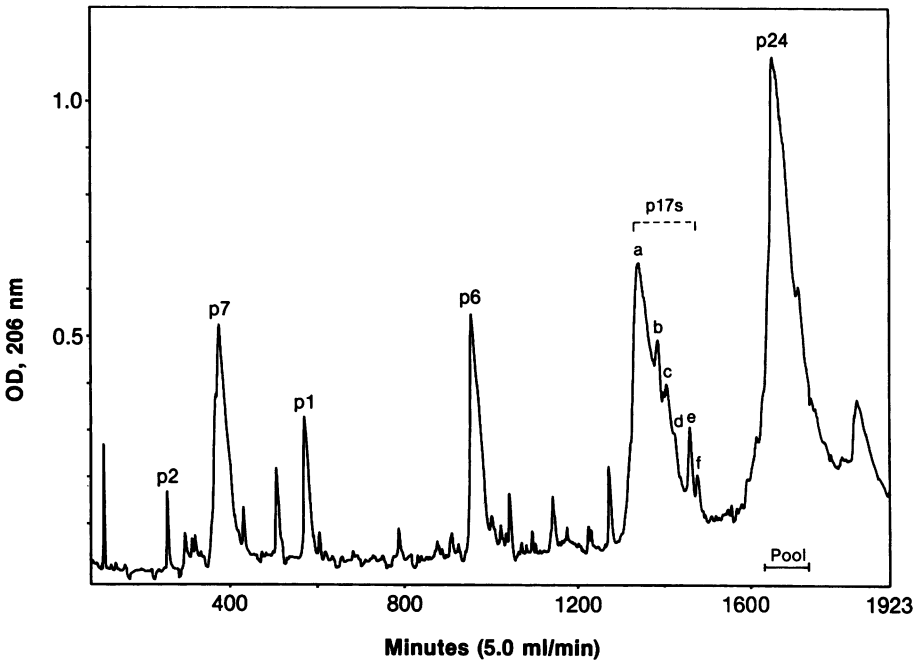

FIG. 1. Protein separation by rp-HPLC. HIV-1_{MN} virions were disrupted, and the viral proteins and peptides were separated by preparative rp-HPLC. Approximately 175 mg of purified HIV-1_{MN} was dissolved and reduced in 40 ml of saturated guanidine hydrochloride containing 2% (vol/vol) 2-mercaptoethanol and injected into a μBondapak C18 column (19 by 150 mm). Proteins and peptides were eluted with a gradient of increasing acetonitrile concentrations at pH 2.0 (0.05% trifluoroacetic acid) and detected by UV absorption at 206 nm. The UV peaks associated with purified viral proteins p17a through p17f, p24, p2, p7, p1, and p6 are indicated. The pool of p24 taken for further analysis is indicated by a bracket. OD, optical density.

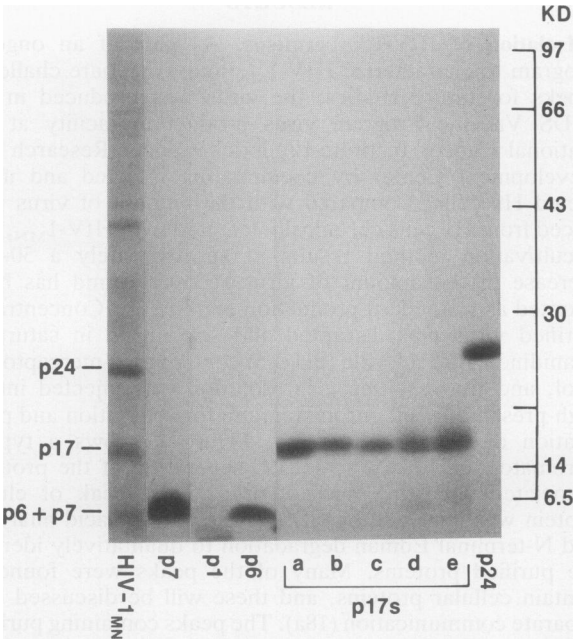

FIG. 2. Coomassie blue-stained gel of purified proteins. HIV-1_{MN} proteins purified as shown in Fig. 1 were analyzed by SDS-PAGE on 6 to 18% polyacrylamide gradient gels and detected by staining with Coomassie brilliant blue R-250 (18). Lane HIV-1_{MN} contained whole disrupted virus, and lanes p7, p1, p6, p24, and p17a through p17e contained approximately 3 to 5 μg of protein purified in the corresponding peak labeled in Fig. 1.

(data not shown). These results (Fig. 1 and 2) show that p1, p2, p6, and p7 are eluted as highly purified proteins in well-separated and nearly symmetrical peaks. p24 is also eluted as a highly purified protein, but the peak is asymmetrical. The trailing edge of the p24 peak contained contaminating proteins (visualized by SDS-PAGE analysis of individual fractions) and was omitted from the pool of highly purified p24 (bracket marked Pool in Fig. 1) taken for further analysis. On the basis of UV absorption, we estimated that the pool contained at least 70% of the total p24 eluted in the asymmetrical peak. In contrast, p17 was eluted from the column as a group of partly separated peaks labeled a through f, as shown in Fig. 1. The protein in each peak migrated as a single band at about 17 kDa in the SDS-PAGE analysis (Fig. 2) and reacted with a polyclonal rabbit antiserum prepared against the protein purified in peak a and also with a mouse monoclonal antibody to HIV-1 p17 (data not shown). Of the total p17 eluted from the column, we estimated that 66% was eluted in peak p17a, 5% was eluted in peak p17b, 9% was eluted in p17c, 9% was eluted in peak p17d, 8% was eluted in peak p17e, and 3% was eluted in peak p17f. To better understand the nature of the apparent chromatographic heterogeneity of the p17 forms and to characterize purified p1, p2, p6, p7, and p24, we subjected the proteins to complete amino acid sequence analysis and examined them for posttranslational modification by mass spectral and chemical methods.

The protein in peak p17a (Fig. 1) was digested with endoproteinase Glu-C, and the peptides were separated by rp-HPLC (Fig. 3). Each purified peptide was analyzed for amino acid composition and amino acid sequence by N-terminal Edman degradation and mass spectrometry. The results are summarized in Fig. 4, in which the determined

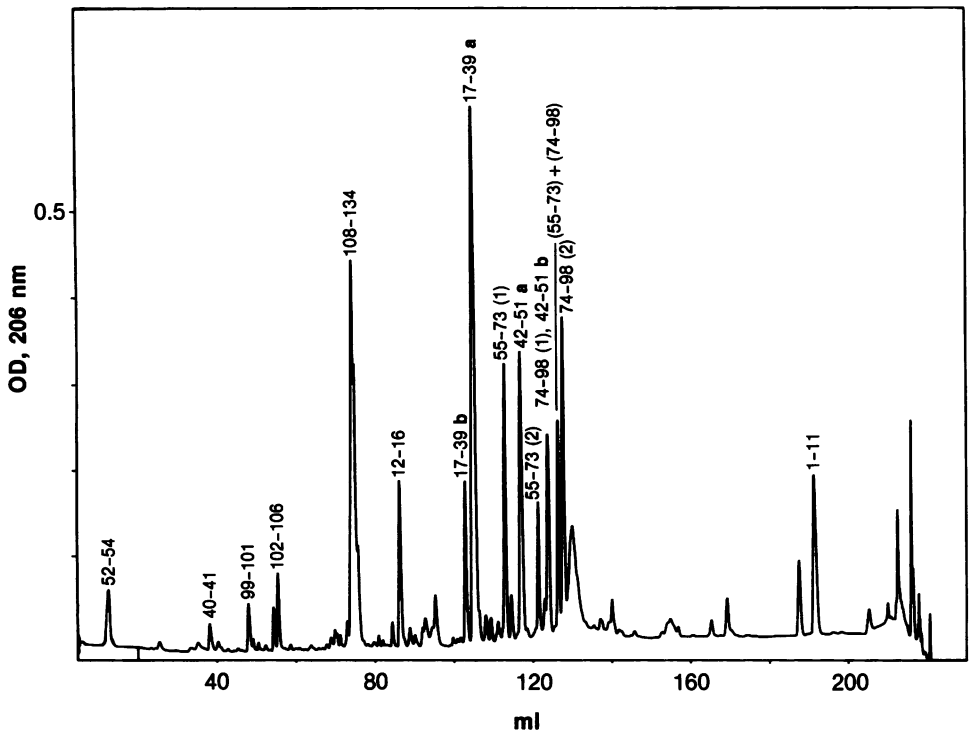

FIG. 3. Separation of peptides resulting from endoproteinase Glu-C digestion of p17a. The protein purified in peak p17a (Fig. 1 and 2) was digested with endoproteinase Glu-C, and the peptides were separated by rp-HPLC. Eluted peptides were detected by UV absorption at 206 nm. The peptide material in each UV peak was identified by amino acid sequencing (see the text), and peptides are numbered according to the positions of their first and last residues in the amino acid sequence alignment shown in Fig. 4. Peptides representing identical regions of p17 (Fig. 4) but differing by amino acid substitutions are distinguished by the letters a and b (see the text). Peptides containing cysteine residues (positions 56 and 86; Fig. 4) were isolated in their monomeric form, indicated by (1), and in their disulfide form, indicated by (2). The peptide designated (55-73)+(74-98) is a heterologous disulfide dimer. When two peptides coeluted in the same peak, both peptides are indicated and separated by a comma. OD, optical density.

amino acid sequences for the peptides are compared with a predicted amino acid sequence based on the nucleotide sequence of codons 2 through 135 of the HIV-1_{MN} proviral *gag* gene (28). The alignment shows that the determined and predicted amino acid sequences are in agreement for 130 of 134 compared positions and prove that the protein in peak p17a (Fig. 1) is derived from the N-terminal end of the Gag precursor. These data also show that the protein in peak p17a is a mixture of at least two closely related p17 forms that differ from each other by substitutions of valine for isoleucine at positions 34 and 46.

The two forms of p17 present in peak p17a (Fig. 1) are revealed by the data in Fig. 3, which shows sets of two peptides that differ from each other by a single amino acid residue. One set is composed of peptides 17-39a and 17-39b (Fig. 3). These peptides have identical amino acid sequences corresponding to positions 17 through 39 in Fig. 4, except that peptide 17-39a has isoleucine at position 34 and peptide 17-39b has valine at position 34. Another set is composed of peptide 42-51a, with a valine at position 46, and peptide 42-51b, with an isoleucine at position 46. The peak designated 74-98(1),42-51b in Fig. 3 is a mixture containing both peptide 74-98 and peptide 42-51b. Edman degradation revealed two residues at each step, confirming the sequences of the two peptides and indicating that peptide 74-98 represented greater than 50% of the total peptide in the mixture. In addition, mass spectral analysis of peak 74-98(1),42-51b (Fig. 3) confirmed the sequence of peptide 42-51b. The determined mass spectra of peptide 42-51b and peptide 42-51a are shown in Fig. 5. The two spectra are identical, except that the masses of ions d_4 to d_9 and a_4 to a_9 in 42-51a (Fig. 5a) are 14 mass units greater than those of the homologous ions in 42-51b (Fig. 5b), confirming the substitution of valine for isoleucine. The amount of peptide 17-39a recovered is about four times the amount of peptide 17-39b recovered (see peak heights in Fig. 3). Similarly, the amount of peptide 42-51a recovered is about four times the amount of peptide 42-51b recovered (Fig. 3). On the basis of these recoveries, we conclude that about 70 to 80% of the protein in p17a (Fig. 1) has isoleucine at position 34 and valine at position 46 and that 20 to 30% of the protein has valine at position 34 and isoleucine at position 46; together, these two forms make up at least 66% of the total p17 eluted from the column. In addition, the determined amino acid sequences of these two forms differ from the predicted sequences at three other positions, at which the proteins have lysine in place of asparagine (position 17) (peptides 17-39a and 17-39b), arginine in place of leucine (position 74) [peptides 74-98(1) and 74-98(2)], and lysine in place of glutamic acid (position 92) (peptide 74-98). Since both the 17-39a and the 17-39b peptides have lysine at position 17 and no other forms of peptide 74-98 were detected, we conclude that both forms of p17 present in the p17a peak (Fig. 1) have lysine at positions 17 and 92 and arginine at position 74.

Peptide 1-11 (Fig. 3) had an amino acid composition consistent with the predicted sequence (based on the DNA

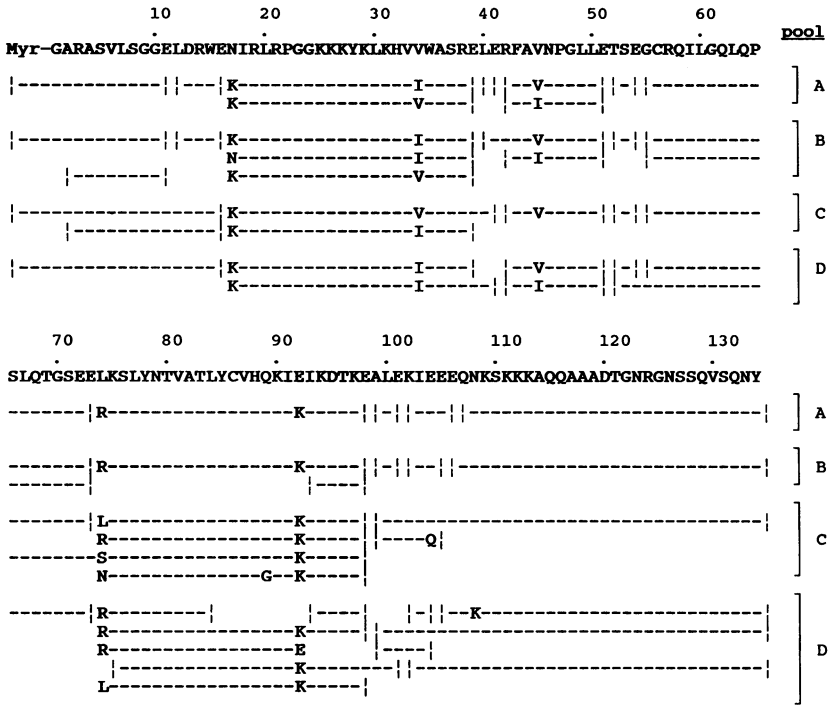

FIG. 4. Amino acid sequence of p17. The amino acid sequences of the chromatographic forms of p17 purified as shown in Fig. 1 (forms p17a to p17d) were determined by analysis of peptides as described in the legend to Fig. 3. The deduced sequence based on the proviral DNA sequence is shown in boldface type in the top line. The determined amino acid sequences of the peptides are shown in pool A for p17a, pool B for p17b, pool C for p17c, and pool D for p17d. Dashes indicate positions at which the deduced and determined sequences agree.

sequence) but was "blocked" to N-terminal Edman degradation. The peptide sequence was determined by mass spectral analysis. All sequence ions derived from the amino terminus of the peptide were 211 mass units greater than the sum of their amino acid residues. In addition, the spectra showed a myristyl-Gly [$CH_3(CH_2)_{12}CONHCH_2CO^+$] ion at 268 mass units and a myristyl [$CH_3(CH_2)_{12}CO^+$] ion at 211 mass units, as previously observed for other myristylated peptides (16, 19). The N-terminal peptide eluted late from the column, consistent with the covalent attachment of a fatty acid, and was immediately preceded by another peptide (unlabeled in Fig. 3). Analysis of this preceding peptide revealed that it was a partial cleavage product consisting of residues 1 to 16 and was also myristylated. No other peptides recovered from the digest contained the N-terminal residue. The molar yields of the N-terminal peptides (1-11 plus 1-16) were equal to the molar yields of the other peptides recovered in Fig. 3. Therefore, we concluded that at least 95% of the protein in peak p17a (Fig. 1), including both forms of p17, contained the N-terminal myristyl moiety. Except for disulfide-bonded peptides (Fig. 3) that may have formed during the digestion, no other posttranslationally modified forms of the protein were detected.

The techniques (endopeptidase digestion, peptide separation and analysis by Edman degradation, and mass spectral analysis) used above for the analysis of protein in peak p17a were used for the analysis of protein purified in peaks p17b, p17c, and p17d (Fig. 1), and the results of these analyses are also summarized in Fig. 4. These results confirmed that the proteins purified in the peaks were viral p17 but also showed that each peak contained a mixture of multiple forms of the protein. Peak p17b contained the forms of p17 identified in peak p17a plus a form of the protein with asparagine at position 17 and isoleucine at position 34. In addition, peptide 93-98 was identified in peak p17b, suggesting that a portion of the protein contained glutamic acid at position 92. Analysis of peak p17c revealed additional forms with leucine, serine, or asparagine at position 74. The leucine at position 74 is in agreement with the nucleotide sequence, but the peptide also contained lysine at position 92, in disagreement with the nucleotide sequence. The recoveries of peptides from the analysis of peak p17c suggested that this peak contained a form of the protein with valine residues at positions 34 and 46, in agreement with the nucleotide sequence. Peak p17c also contained forms of the protein with glycine at position 89 and glutamine at position 104. Peak p17d contained additional forms of the protein with lysine at position 108 and possibly glutamic acid at positions 74, 84, and 92 (also indicated in p17b).

The total number of amino acid sequence variants of p17 present in the virus population cannot be estimated from these results, because some variants may be present in amounts below our limits of detection. However, the analysis of the various chromatographic forms of p17 revealed at least eight different amino acid sequence variants of the protein. In addition, the proteins in peaks p17e and p17f were not analyzed but probably contained additional variant forms. Nevertheless, at least one-half of the total p17 is contributed by the major form in peak p17a, and about one-fifth of the total p17 is contributed by the minor form in p17a. None of the forms identified were found to be in complete agreement with the protein sequence deduced from the proviral DNA sequence of HIV-1_{MN}. At least 98% of the protein analyzed contained the N-terminal myristyl posttranslational modification, but peptides that lacked the mod-

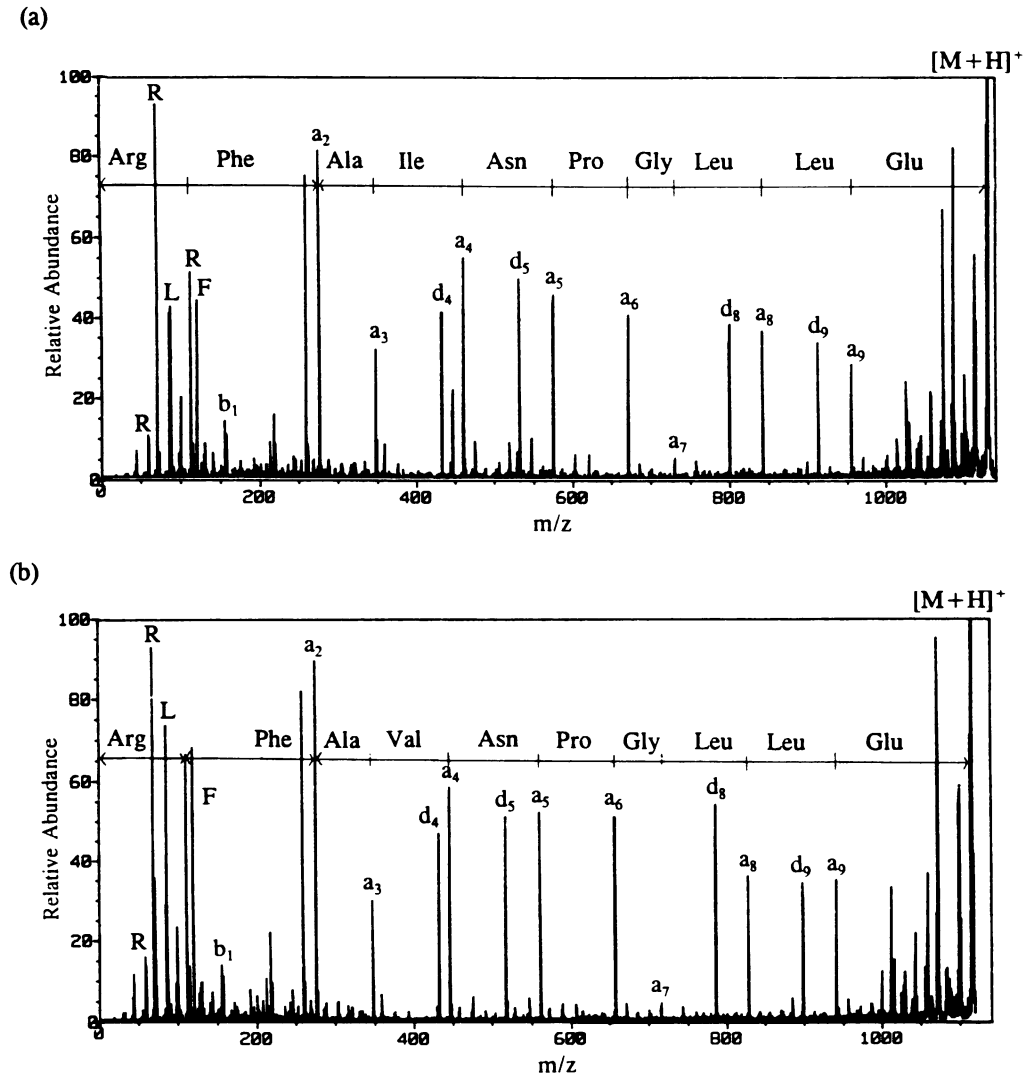

FIG. 5. Tandem mass spectra of peptides 42-51a and 42-51b (Fig. 3). Both peptides represent positions 42 to 51 in Fig. 4; residue 45 is isoleucine in 42-51a (spectrum a) and valine in 42-51b (spectrum b). The fragment ions that contain residue 45 are 14 mass units heavier in spectrum a than in spectrum b because of the presence of the isoleucine in place of a valine.

ified N-terminal glycine residue were noted in the analysis of p17b and p17c.

ESI-MS of p17. The protein purified in peak p17a (Fig. 1) was analyzed by ESI-MS. This is a highly accurate method for determining the molecular weight of a protein and is useful for verifying amino acid compositions of proteins and for detecting the presence of posttranslational modifications present before proteolysis (26). Figure 6 shows the ESI-MS spectrum obtained from the analysis of p17a (Fig. 1). Analysis of the ion groups with different charged states revealed a molecular weight of 15,269.3 mass units for the protein. This value is in near-perfect agreement with the molecular weight of 15,266.3 mass units calculated from the determined structures of both forms of p17 (Fig. 4) present in peak p17a (Fig. 1). Close agreement between the determined and calculated molecular weights strongly suggests that the p17 analyzed here (p17a) is posttranslationally modified by the addition of an N-terminal myristyl group but contains no other modifications. However, this conclusion cannot be extended to the other chromatographic forms of p17 (p17b to p17f; Fig. 1), since they were not analyzed by ESI-MS. Phosphorylated forms of p17 have been detected by ^{32}P labeling of HIV-1_{LAV} viral proteins (27), and it is quite possible that one or more of the minor chromatographic forms of p17 contain phosphorylated residues. The finding that the predominant forms of p17 are not phosphorylated is consistent with an earlier analysis of p17 purified from HIV-1_{IIIB} in which we did not detect the presence of phosphorylated amino acids (15). Our previous analysis was done with nonradioactive methods and would not have detected phosphoproteins present in <20% of the total protein analyzed.

A pool of highly purified p24 (marked in Fig. 1) containing at least 70% of the total p24 eluted from the column was digested with endoproteinase Glu-C or trypsin. For each digest, the resulting peptides were separated by rp-HPLC (data not shown) and analyzed as described for the various p17 forms. The determined amino acid sequences were aligned with the amino acid sequence predicted by the proviral DNA sequence (Fig. 7). Only the relevant and

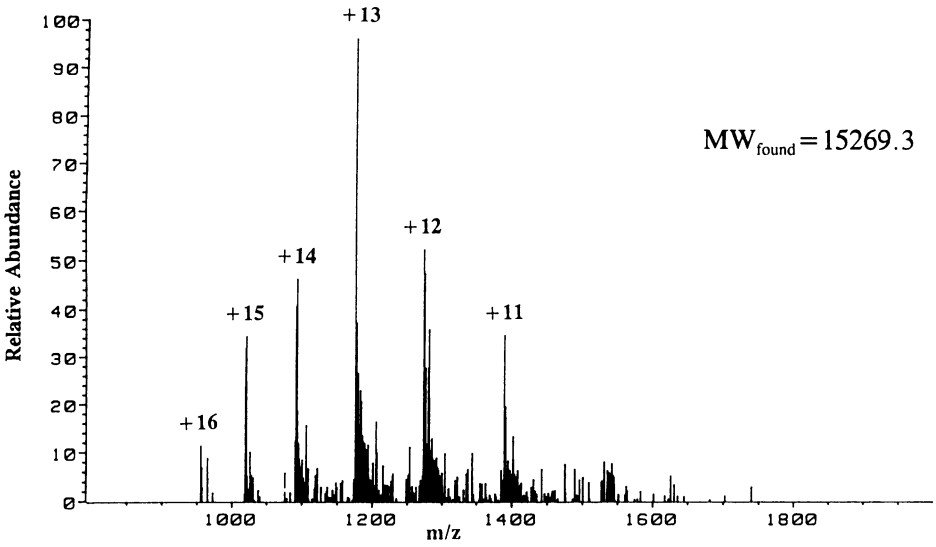

FIG. 6. ESI mass spectrum of p17a (Fig. 1). Each peak in the envelope is a molecular ion with the indicated charge state. A mass spectrometer measures mass/charge; hence, the molecular weight of a large analyte, such as p17, can be determined if the analyte has a high charge state (26). These data yield a determined molecular weight of 15,269.3 for p17a (Fig. 1).

nonredundant peptides are shown. These analyses revealed two amino acid sequence variants of p24. The predominant form represented about 70% of the total protein analyzed and contained alanine at position 86; the other form represented about 30% of the total protein analyzed and contained valine at position 86. Both forms differed from the predicted amino acid sequence at three other positions: position 7, at which the protein contained glutamine in place of glutamic acid; position 184, at which the protein contained tryptophan in place of arginine; and position 185, at which the protein contained methionine in place of threonine.

Protein in the p24 pool (Fig. 1) was also analyzed by ESI-MS, and a molecular weight of 25,752 ± 50 was determined (data not shown). The calculated weights for the two forms of p24 (Fig. 7) were 25,551 (alanine at position 86) and 25,579 (valine at position 86). Thus, the measured molecular weight is about 200 greater than the calculated molecular weight. Previous reports have shown that HIV-1 p24 (from strain IIIB or BRU) contains phosphoserine and phosphothreonine (15, 27) and exhibits multiple isoelectric forms containing phosphate (24). The ESI-MS data do not directly determine phosphates; however, the results of the analysis are consistent with the previous proposals that the p24 protein is phosphorylated. The observed variation in molecular weight (±50) is greater than the precision of the ESI-MS method (26) and is in part due to the sequence heterogeneity in the sample but also could suggest that the sample contains a mixture of diphosporylated and triphosphorylated proteins. Unfortunately, mass spectral analysis of the proteolytic peptides derived from p24 (as in Fig. 7) failed to reveal any evidence of posttranslational modifications, including phosphorylations. It seems possible that putative phosphate groups were lost through hydrolysis during the enzymatic digestion of the protein or separation of the peptides.

The other Gag proteins purified in Fig. 1 (p1, p2, p6, and p7) were also analyzed by Edman degradation of purified peptides and mass spectral analysis of peptides (data not shown). The results showed that the amino acid sequences

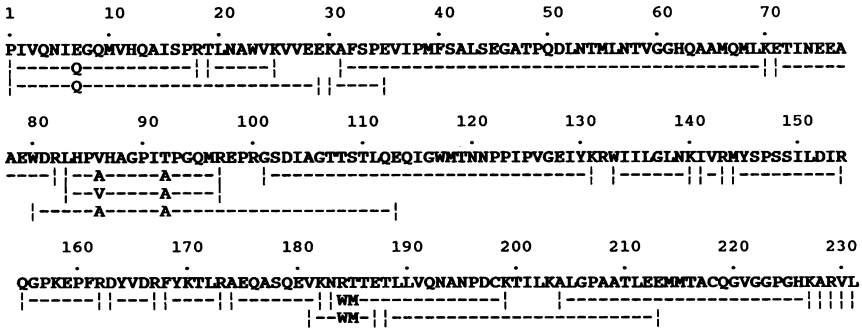

FIG. 7. Amino acid sequence of p24. Peptides derived from trypsin and endoproteinase Glu-C proteolytic digestions of p24 (Fig. 1) were isolated by rp-HPLC and sequenced by Edman degradation and mass spectral analysis. The determined amino acid sequences are aligned with the sequence deduced from the proviral DNA sequence shown in boldface type in the top line. Dashes indicate locations at which the determined sequences agree with the deduced sequence. Two variants of peptide 83-96 were found (amino acid sequence variations at position 86). The most abundant form of peptide 83-96 (approximately 70% of the total) contained alanine at position 86.

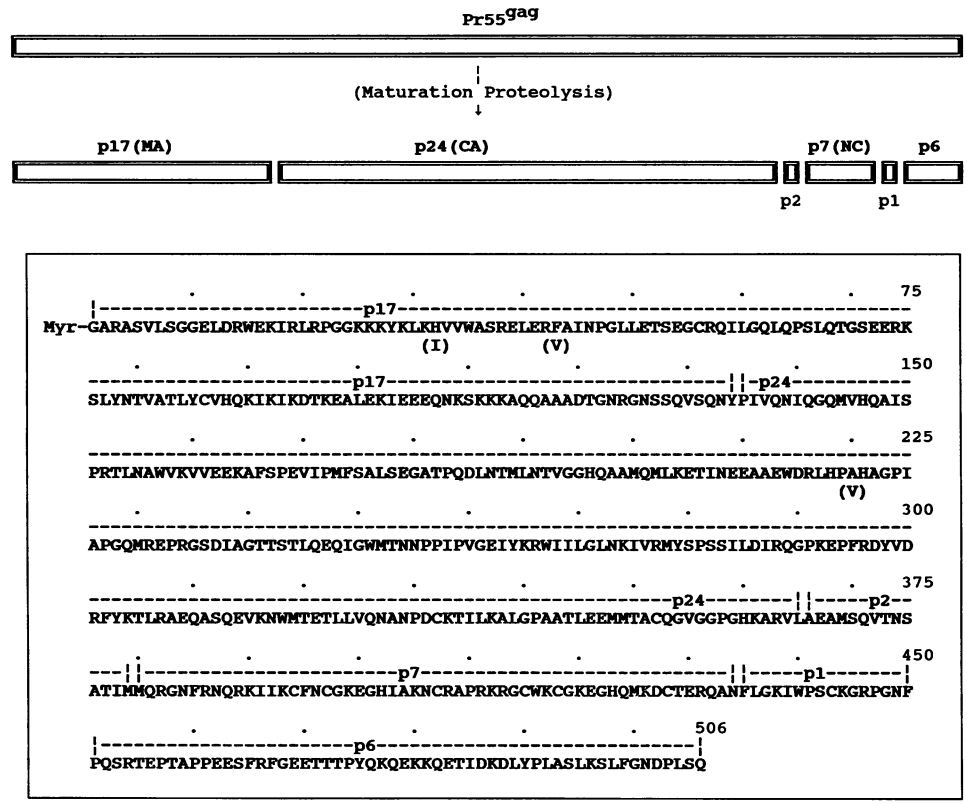

FIG. 8. (Top) Schematic diagram showing the maturation proteolytic processing of the HIV-1_{MN} Gag precursor to the Gag proteins (p17, p24, p2, p7, p1, and p6) found in the mature virus. (Bottom) Amino acid sequence of a functional Gag precursor from the most infectious replicative variants of HIV-1_{MN} grown in H9 cells. Amino acid residues in parentheses indicate substitutions that distinguish another competent Gag precursor also present in the infectious replicative variants.

of purified p2, p6, and p7 were identical to the amino acid sequences predicted by the proviral DNA sequence, and there was no evidence of amino acid sequence variants or posttranslational modifications for these proteins. Purified p1 (Fig. 1) was found to have the amino acid sequence FLGKIWPSCKGRPGNF, which differs from the predicted sequence at position 13, at which the p1 protein contains proline in place of a deletion, indicated by the DNA sequence, and at position 14, at which the p1 protein contains glycine in place of arginine. No posttranslational modifications or sequence variants of p1 were detected.

In summary, the analysis of the purified Gag proteins shows the complete amino acid sequences and maturation proteolytic cleavage sites for the Gag precursors that constitute the major portion of HIV-1_{MN} produced by infection and replication in H9 cells (Fig. 8). The amino acid sequence of the most abundant form of the Gag precursor (approximately 50% of the total) is shown in Fig. 8, and that of the next most abundant form (approximately 20% of the total) is also indicated. The schematic diagram showing the maturation proteolytic cleavage sites of the Gag precursors (Fig. 8) indicates the cleavage sites for generating the proteins described here (p17, p24, p2, p7, p1, and p6). The cleavage sites shown in Fig. 8 are highly efficient sites, since no appreciable partial cleavage products were identified. However, other peptide fragments of the Gag proteins indicating less efficient cleavage sites were also noted during the course of this work. These included fragments of p2, p1, and p6, indicating cleavages between residues 369 and 370, 439 and 440, and 490 and 491 in the diagram shown in Fig. 8.

DISCUSSION

Products of the retroviral *gag* gene perform highly complex orchestrated tasks during the assembly, budding, maturation, and infection stages of the viral replication cycle. During viral assembly, the proteins form membrane associations (12, 30, 32) and self-associations that ultimately result in budding of an immature virion from the infected cell, and point mutations or small deletions in the *gag* gene can exert a dominant repressive effect on viral production (35). Gag precursors also function during viral assembly to selectively bind and package two plus strands of genomic RNA (10, 11). The assembled immature virion contains approximately 2,000 to 3,000 Gag precursors (3, 20) and becomes infectious after activation of the viral protease (21), proteolytic cleavage of the Gag and Gag-Pol precursors, and rearrangement of the products to form the mature virion. In the mature virion, the Gag proteins associate to form specifically defined structures (9) that are probably necessary for the infection process.

The amino acid sequences of Gag precursors have been deduced from the DNA sequences of proviruses derived from numerous strains of HIV-1 (28). Many of these proviruses are known to be infectious; however, it is difficult or impossible to deduce the exact amino acid sequences of highly efficient Gag precursors from the study of proviral DNA sequences. The most direct way to determine structures of Gag precursors that function efficiently in viral assembly is to determine exact amino acid sequences of Gag proteins incorporated into highly replicative viruses. H9

cells chronically infected with HIV-1_{MN} contain many integrated proviruses (31) but do not shed large amounts of virions into the culture medium (2). However, when chronically infected H9 cells are continuously cocultivated with uninfected H9 cells, the amounts of virus shed into the medium increase by at least 50-fold. This biological amplification selects for the most infectious replicative viruses.

Here we have purified mature Gag proteins directly from large-scale production lots of HIV-1_{MN} and determined their complete amino acid sequences and posttranslational modifications, including the maturation proteolytic cleavage sites in the precursors. We have used complimentary, but independent, methods (mass spectrometry and automated Edman degradation plus amino acid analysis) (4, 8) to determine the amino acid sequences and posttranslational modifications and have confirmed the results by independently determining the exact molecular weights of the purified proteins by using ESI-MS. This combined approach creates a large body of overlapping and redundant data that results in a very high level of confidence in the final determined structures.

The results show that during maturation, the HIV-1 Gag precursors are cleaved to yield six mature Gag products, as indicated in Fig. 8, and that the order of cleavage products in the precursors is p17-p24-p2-p7-p1-p6. The cleavage sites and number of proteolytic products are in agreement with previous findings for HIV-1_{IIIB} (15) and are also highly homologous to the cleavage sites and number of products previously identified for simian immunodeficiency virus (14) and related viruses, such as HIV-2 (18). The proteolytic processing scheme proposed here reveals five cleavage sites and differs from other proposals indicating fewer cleavage sites (27, 36). All six cleavage products were isolated from HIV-1_{MN} (Fig. 1) and, on the basis of the amounts of the proteins, we estimate that approximately equal molar amounts of each Gag protein were recovered from the virus. These findings suggest that cleavage at each site occurs by an efficient process and are in agreement with the equal molar recoveries of Gag precursor cleavage products previously reported for other retroviruses (14, 17).

Each Gag protein was recovered as a single chromatographic entity, with the exception of p17, for which at least six chromatographically distinct forms were noted (peaks p17a to p17f in Fig. 1). Analysis of p17a to p17d revealed that a minimum of eight different amino acid sequence variants of this protein were present in the total HIV-1_{MN} population (Fig. 4). Quantitative recoveries indicated that a single amino acid sequence variant of p17 (Fig. 4) contributed at least 50% to the total p17 recovered from the virus and that another variant (also identified in Fig. 4) contributed an additional 20 to 30% to the total. The remainder of the total p17 was divided among the other amino acid sequence variants of p17 (Fig. 4), with no single variant representing more than about 5 to 8% of the total. Analysis of p24 revealed two amino acid sequence variants (Fig. 7). The most abundant variant contributed at least 50% to the total p24 recovered from the virus, and the other variant contributed at least 20% to the total (30% of the total p24 was not analyzed). No other amino acid sequence variants were revealed by analysis of the total protein recovered for each of the other Gag proteins (p1, p2, p6, and p7; Fig. 1). On the basis of quantitative recoveries and amino acid sequences, we suggest that approximately 50 to 60% of the total Gag proteins recovered are proteolytic cleavage fragments of a single Gag precursor and that an additional 20 to 30% of the total Gag proteins may originate from another Gag precursor, as indicated in Fig. 8. The sequence heterogeneity probably results from the expression of several different proviruses integrated in the chronically infected cell population. The total number of different viruses expressed in the mixture cannot be estimated, but the most replicative viruses will be expanded in the cocultivation procedure. The data provided here help characterize the more replicative viruses and show the amino acid sequences of Gag precursors that function efficiently in the assembly and budding process.

In addition to amino acid sequences, posttranslational modifications can greatly influence biological properties and must be considered when describing the molecular compositions of proteins. A previous quantitative analysis of the IIIB (15) and LAV (27) strains of HIV-1 suggested that most MA proteins in the mature viruses were myristylated. Here we analyzed five chromatographic variants of the MA protein produced by the MN strain and showed that each variant was myristylated. Together, these variants represent the vast majority of the total MA protein present in the bulk virus; therefore, these data suggest that the most prevalent forms of the viruses produced from the MN strain utilize myristylated Gag precursors during viral assembly. The importance of the myristylation modification (30, 33) is underscored by the recent finding that analogs of myristic acid can inhibit virus production (5, 6). A search for other possible posttranslational modifications was conducted by comparing experimentally determined molecular weights (mass spectrometry) and calculated molecular weights (sequencing) of the purified Gag proteins. The observed molecular weight of the protein in the pool of p24 analyzed was 200 $\pm$ 50 weight units higher than the calculated molecular weight, indicating the presence of posttranslational modifications. The nature of the modifications was not directly determined; however, these data are consistent with the previous observations that p24 contains phosphoserine and phosphothreonine (15, 27).

The results presented here were obtained from a single production lot of HIV-1_{MN} and clearly show amino acid sequence heterogeneity in the isolated virus. The methods used here can be applied to other production lots to help determine sequence variations in the virus that may arise through continued culturing. Other lots have been analyzed by rp-HPLC and show chromatographic evidence of heterogeneity. However, a more refined chromatographic procedure must be developed to analyze and compare the sequence heterogeneity in each lot.

ACKNOWLEDGMENTS

We thank Clara M. Dinterman and Patricia Coulter Grove for editorial and clerical assistance.

This research was sponsored, at least in part, by the National Cancer Institute under contract N01-CO-74102 with Program Resources, Incorporated/DynCorp. This work was also supported, in part, by grant BBS 87-14238 from the National Science Foundation. Mass spectral measurements were performed at the Structural Biochemistry Center, an NSF-supported Biological Instrument Center at the University of Maryland, Baltimore County.

REFERENCES

1. **Barré-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vézinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science **220**:868–871.
2. **Bess, J. W., Jr. (Program Resources, Incorporated/Dyn Corp,**

National Cancer Institute-Frederick Cancer Research and Development Center). 1991. Personal communication.
3. **Bess, J. W., Jr., L. E. Henderson, and L. O. Arthur.** Unpublished data.
4. **Biemann, K., and H. A. Scoble.** 1987. Characterization by tandem mass spectrometry of structural modifications in proteins. Science **237:**992–998.
5. **Bryant, M. L., R. O. Heuckeroth, J. T. Kimata, L. Ratner, and J. I. Gordon.** 1989. Replication of human immunodeficiency virus 1 and Moloney murine leukemia virus is inhibited by different heteroatom-containing analogs of myristic acid. Proc. Natl. Acad. Sci. USA **86:**8655–8659.
6. **Bryant, M. L., L. Ratner, R. J. Duronio, N. S. Kishore, B. Devadas, S. P. Adams, and J. I. Gordon.** 1991. Incorporation of 12-methoxydodecanoate into the human immunodeficiency virus 1 *gag* polyprotein precursor inhibits its proteolytic processing and virus production in a chronically infected human lymphoid cell line. Proc. Natl. Acad. Sci. USA **88:**2055–2059.
7. **Devash, Y., T. J. Matthews, J. E. Drummond, K. Javaherian, D. J. Waters, L. O. Arthur, W. A. Blattner, and J. R. Rusche.** 1990. C-terminal fragments of gp120 and synthetic peptides from five HTLV-III strains: prevalence of antibodies to the HTLV-III-MN isolate in infected individuals. AIDS Res. Hum. Retroviruses **6:**307–316.
8. **Fenselau, C.** 1991. Beyond gene sequencing: analysis of protein structure with mass spectrometry. Annu. Rev. Biophys. Biophys. Chem. **20:**205–220.
9. **Gelderbloom, H. R., M. Özel, and G. Pauli.** 1989. Morphogenesis and morphology of HIV, structure-function relations. Arch. Virol. **106:**1–13.
10. **Gorelick, R., L. Henderson, J. Hanser, and A. Rein.** 1988. Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a "zinc finger-like" protein sequence. Proc. Natl. Acad. Sci. USA **85:**8420–8424.
11. **Gorelick, R. J., S. M. Nigida, Jr., J. W. Bess, Jr., L. O. Arthur, L. E. Henderson, and A. Rein.** 1990. Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. J. Virol. **64:**3207–3211.
12. **Göttlinger, H. G., J. G. Sodroski, and W. A. Haseltine.** 1989. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. USA **86:**5781–5785.
13. **Gurgo, C., H.-G. Guo, G. Franchini, A. Aldovini, E. Collalti, K. Farrell, F. Wong-Staal, R. C. Gallo, and M. S. Reitz, Jr.** 1988. Envelope sequences of two new United States HIV-1 isolates. Virology **164:**531–536.
14. **Henderson, L. E., R. E. Benveniste, R. Sowder, T. D. Copeland, A. M. Schultz, and S. Oroszlan.** 1988. Molecular characterization of *gag* proteins from simian immunodeficiency virus (SIV_{Mne}). J. Virol. **62:**2587–2595.
15. **Henderson, L. E., T. D. Copeland, R. C. Sowder, A. M. Schultz, and S. Oroszlan.** 1988. Analysis of proteins and peptides purified from sucrose gradient banded HTLV-III, p. 135–147. *In* D. Bolognesi (ed.), Human retroviruses, cancer, and AIDS: approaches to prevention and therapy. Alan R. Liss, Inc., New York.
16. **Henderson, L. E., H. C. Krutzsch, and S. Oroszlan.** 1983. Myristyl amino-terminal acylation of murine retrovirus proteins: an unusual post-translational protein modification. Proc. Natl. Acad. Sci. USA **80:**339–343.
17. **Henderson, L. E., R. Sowder, T. D. Copeland, G. Smythers, and S. Oroszlan.** 1984. Quantitative separation of murine leukemia virus proteins by reversed-phase high-pressure liquid chromatography reveals newly described *gag* and *env* cleavage products. J. Virol. **52:**492–500.
18. **Henderson, L. E., R. C. Sowder, T. D. Copeland, S. Oroszlan, and R. E. Benveniste.** 1990. *Gag* precursors of HIV and SIV are cleaved into six proteins found in the mature virions. J. Med. Primatol. **19:**411–419.
18a. **Henderson, L. E., R. C. Sowder II, D. G. Johnson, J. W. Bess, Jr., and L. O. Arthur.** Unpublished data.
19. **Hizi, A., L. E. Henderson, T. D. Copeland, R. C. Sowder, H. C. Krutzsch, and S. Oroszlan.** 1989. Analysis of *gag* proteins from mouse mammary tumor virus. J. Virol. **63:**2543–2549.
20. **Karpel, R. L., L. E. Henderson, and S. Oroszlan.** 1987. Interactions of retroviral structural proteins with single-stranded nucleic acids. J. Biol. Chem. **262:**4961–4967.
21. **Kohl, N. E., E. A. Emini, W. E. Schleif, L. J. Davis, J. C. Heimbach, R. A. F. Dixon, E. M. Scolnick, and I. S. Sigal.** 1988. Active human immunodeficiency virus protease is required for viral infectivity. Proc. Natl. Acad. Sci. USA **85:**4686–4690.
22. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
23. **LaRosa, G. J., J. P. Davide, K. Weinhold, J. A. Waterbury, A. T. Profy, J. A. Lewis, A. J. Langlois, G. R. Dreesman, R. N. Boswell, P. Shadduck, L. H. Holley, M. Karplus, D. P. Bolognesi, T. J. Matthews, E. A. Emini, and S. D. Putney.** 1990. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. Science **249:**932–935.
24. **Laurent, A. G., B. Krust, M.-A. Rey, L. Montagnier, and A. G. Hovanessian.** 1989. Cell surface expression of several species of human immunodeficiency virus type 1 major core protein. J. Virol. **63:**4074–4078.
25. **Levy, J. A., A. D. Hoffman, S. M. Kramer, J. A. Landis, J. M. Shimabukuro, and L. S. Oshiro.** 1984. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. Science **225:**840–842.
26. **Loo, J. A., C. G. Edmonds, and R. D. Smith.** 1990. Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. Science **248:**201–204.
27. **Mervis, R. J., N. Ahmad, E. P. Lillehoj, M. G. Raum, F. H. R. Salazar, H. W. Chan, and S. Venkatesan.** 1988. The *gag* gene products of human immunodeficiency virus type 1: alignment within the *gag* open reading frame, identification of posttranslational modifications, and evidence for alternative *gag* precursors. J. Virol. **62:**3993–4002.
28. **Myers, G., J. A. Berzofsky, B. Korber, R. F. Smith, and G. N. Pavlakis (ed.).** 1991. Human retroviruses and AIDS 1991, a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, N.Mex.
29. **Popovic, M., M. G. Sarngadharan, E. Read, and R. C. Gallo.** 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science **224:**497–500.
30. **Rein, A., M. R. McClure, N. R. Rice, R. B. Luftig, and A. M. Schultz.** 1986. Myristylation site in $Pr65^{gag}$ is essential for virus particle formation by Moloney murine leukemia virus. Proc. Natl. Acad. Sci. USA **83:**7246–7250.
31. **Reitz, M. (National Cancer Institute).** 1991. Personal communication.
32. **Rhee, S. S., and E. J. Hunter.** 1990. Structural role of the matrix protein of type D retroviruses in Gag polyprotein stability and capsid assembly. J. Virol. **64:**4383–4389.
33. **Schultz, A., and A. Rein.** 1989. Unmyristylated Moloney murine leukemia virus $Pr65^{gag}$ is excluded from virus assembly and maturation events. J. Virol. **63:**2370–2373.
34. **Toplin, I., and P. Sottong.** 1972. Large-volume purification of tumor viruses by use of zonal centrifuges. Appl. Microbiol. **23:**1010–1014.
35. **Trono, D., M. B. Feinberg, and D. Baltimore.** 1989. HIV-1 Gag mutants can dominantly interfere with the replication of the wild-type virus. Cell **59:**113–120.
36. **Veronese, F. D., T. D. Copeland, S. Oroszlan, R. C. Gallo, and M. G. Sarngadharan.** 1988. Biochemical and immunological analysis of human immunodeficiency virus *gag* gene products p17 and p24. J. Virol. **62:**795–801.