JOURNAL OF VIROLOGY, Jan. 1989, p. 111–121
0022-538X/89/010111-11$02.00/0

Vol. 63, No. 1

Mutational Analysis of Human Immunodeficiency Virus Type 1 Protease Suggests Functional Homology with Aspartic Proteinases

DANIEL D. LOEB,[1] CLYDE A. HUTCHISON III,[1] MARSHALL H. EDGELL,[1] WILLIAM G. FARMERIE,[2] AND RONALD SWANSTROM[2]*

Department of Microbiology and Immunology, Curriculum in Genetics, Program in Molecular Biology and Biotechnology,[1] and Department of Biochemistry and Lineberger Cancer Research Center,[2] University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Received 11 July 1988/Accepted 8 September 1988

Processing of the retroviral *gag* and *pol* gene products is mediated by a viral protease. Bacterial expression systems have been developed which permit genetic analysis of the human immunodeficiency virus type 1 protease as measured by cleavage of the *pol* protein precursor. Deletion analysis of the *pol* reading frame locates the sequences required to encode a protein with appropriate proteolytic activity near the left end of the *pol* reading frame but largely outside the *gag-pol* overlap region, which is at the extreme left end of *pol*. Most missense mutations within an 11-amino-acid domain highly conserved among retroviral proteases and with sequence similarity to the active site of aspartic proteinases abolish appropriate processing, suggesting that the retrovirus proteases share a catalytic mechanism with aspartic proteinases. Substitution of the amino acids flanking the scissile bond at three of the processing sites encoded by *pol* demonstrates distinct sequence requirements for cleavage at these different sites. The inclusion of a charged amino acid at the processing site blocks cleavage. A subset of these substitutions also inhibits processing at the nonmutated sites.

The *pol* gene of retroviruses generally encodes three distinct proteins: a protease (PR) encoded at the 5′ end of the gene; the reverse transcriptase-RNase H (RT) in the center of the gene; and the integration protein (IN) encoded at the 3′ end of the gene. Expression of *pol* is controlled at the level of translation. The *pol* gene is expressed as a fusion protein with *gag* (encoded upstream of *pol*) after either termination suppression (when *gag* and *pol* are in the same reading frame [46]) or a frameshifting event (when *gag* and *pol* are in different partially overlapping reading frames [18–20, 32, 47]). The *gag* and *pol* gene products are processed out of the *gag* and *gag-pol* precursors by the *pol*-encoded PR (7, 43, 44, 48). Processing is an obligatory step in the virus life cycle, since mutations in the PR domain result in the production of noninfectious virus with unprocessed *gag* and *gag-pol* precursor proteins (3, 22, 23).

Genetic evidence suggests that in the human retrovirus human immunodeficiency virus type 1 (HIV-1), the etiologic agent of acquired immunodeficiency syndrome, an analogous PR activity is encoded at the 5′ end of the *pol* reading frame (8, 23, 26, 33), and a small domain from this region of the genome has been expressed in *Escherichia coli* to give a PR capable of cleaving the *gag* or *pol* precursor (5, 9). This region encodes amino acids that are conserved among retroviruses, and some of these amino acids share sequence identity with the active site of aspartic proteinases (40). The sequence of the amino terminus of an 11-kilodalton (kDa) virion protein (p11), thought to be the PR, has been determined previously (30). The other proteins encoded within the HIV-1 *pol* gene are two forms of the RT that share a common amino terminus but differ in molecular mass (64 and 51 kDa; p64 and p51, respectively [6, 29]), a protein of 15 kDa that has RNase H activity (12), and a protein of molecular mass 34 kDa (p34) which is the presumed viral integration protein (29). For these proteins to be generated, at least four processing events must occur within the *pol* portion of the *gag-pol* precursor (Fig. 1A).

We and others have developed bacterial expression systems in which the HIV-1 *pol* gene is expressed and its primary translation product is processed to give proteins similar in size to those found in virus particles (8, 13, 28, 30). Analogous expression systems have been described for the PR of the avian retrovirus Rous sarcoma virus, in which the PR is encoded by *gag* (25, 31). The appearance of these viral proteins is dependent on coexpression of the presumed viral PR domain, indicating that processing in this system is mediated by the *pol*-encoded protease. The use of heterologous expression systems provides a simple and quick assay for HIV-1 PR function and for processing of the *pol* precursor. Using this bacterial expression system, we carried out a series of mutational analyses to study processing of the HIV-1 *pol* gene product.

MATERIALS AND METHODS

Plasmid constructions. The 3.7-kilobase-pair *Bgl*II-*Sal*I fragment from the cloned HIV-1 provirus HXB2 (36) was subcloned into the *Bam*HI-*Sal*I sites in the polylinker region of pIBI20 to make pART1. This fragment of the viral genome contains the entire *pol* gene with the exception of the first four codons. In the pIBI20 plasmid, the insert is under the inducible control of the *lac* promoter-operator. The fused *Bam*HI-*Bgl*II site at the 5′ end of the *pol* gene insert in pART1 was converted into a *Bgl*II site (to make pART2) by a single point change (G to A) which did not change the protein-coding capacity of pART1. The plasmids pBRT1prt$^+$ and pBRT3prt$^-$ were previously described (8).

Deletion clone construction. The method of Henikoff (15) as modified by Hoheisel and Pohl (16) was used to construct unidirectional deletions in the 5′ end of the *pol* gene. The plasmid pART1 was linearized by digestion with *Sst*I and *Sma*I, two adjacent sites in the polylinker 5′ to the *pol* gene insert. The *Sma*I terminus was susceptible to digestion with exonuclease III and nuclease S1, which resulted in deletions

* Corresponding author.

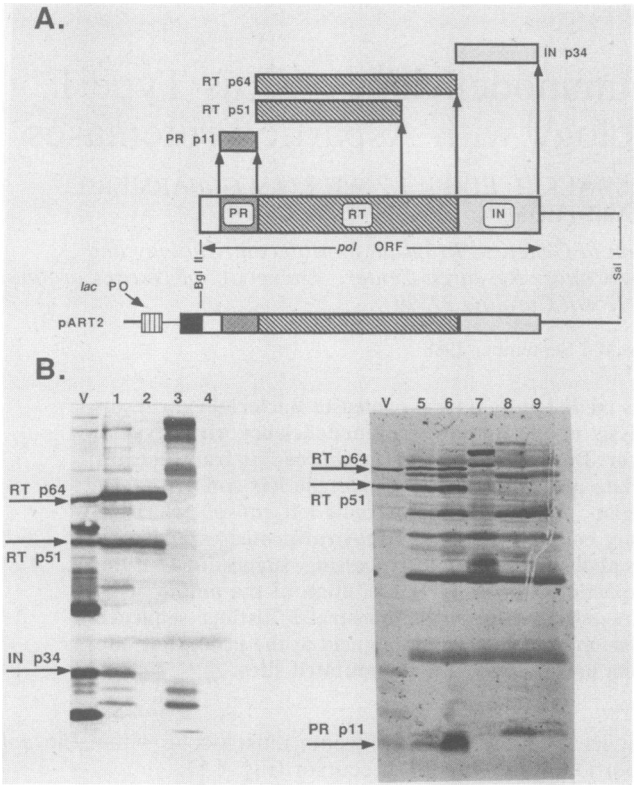

FIG. 1. Expression of processed *pol* gene products in *E. coli*. (A) The large box represents the *Bgl*II-*Sal*I fragment from the cloned HIV-1 provirus HXB2 (36), which contains the *pol* open reading frame (ORF) with domains (PR, RT, and IN) as indicated. Above this box are illustrated the four mature products derived from the *pol* ORF. The PR is labeled PR p11; the RT is labeled RT p64 and RT p51; and the IN is labeled IN p34. The vertical arrows (with the exception of the one at the right end of the IN domain) represent protein-processing sites in the *pol* precursor. Below the large box is a drawing illustrating features of the pART2 *pol* expression plasmid: ▥, *lac* promoter-operator (*lac* PO); ■, protein-coding sequence contributed by the vector (23 codons of *lacZ*-polylinker region from pIBI20). (B) Two Western blots demonstrating the ability of bacteria containing pART2 to direct the synthesis of the four mature *pol* gene products. The left panel is a Western blot of proteins separated by electrophoresis in a 10% polyacrylamide gel, transferred to nitrocellulose paper, and then stained with a 1:500 dilution of an HIV-1-seropositive human antiserum. Lanes: V, lysate of pelleted HIV-1; 1, lysate of bacteria expressing pART2; 2, lysate of bacteria expressing pBRT1prt$^+$ which is truncated in the IN domain (8); 3, lysate of bacteria expressing a derivative of pART1 (pART1E2) which has a frame shift in the polylinker region upstream of the 5′ end of the *pol* insert (to identify internal initiations); 4, lysate of bacteria containing pIBI20 (the parent vector). The right panel is a Western blot of proteins separated in a 15% polyacrylamide gel, transferred to nitrocellulose paper, and stained with a 1:1,000 dilution of a polyclonal rabbit antiserum directed against the C terminus of the HIV-1 PR and the N terminus of the RT (2). Lanes: V, lysate of pelleted HIV-1; 5, lysate of bacteria expressing pART2; 6, lysate of bacteria expressing pBRT1prt$^+$; 7, lysate of bacteria expressing pBRT3prt$^-$ which is truncated in both the PR and IN domains (8); 8, same as lane 3; 9, same as lane 4.

of various lengths extending into the HIV-1 *pol* insert. Digestion products were circularized with ligase and recovered as deletion clones by transformation of *E. coli* JM101.

Design and synthesis of mutagenic oligonucleotides. To introduce random missense mutations into selected domains of the PR-coding sequence, two adjacent 39- and 40-base oligonucleotides were synthesized under conditions such that an average of 1.5 mutations or misincorporations per oligonucleotide were achieved as described by Hutchison et al. (17). To make random changes in the two codons surrounding the scissile bonds of three different cleavage sites in the *pol* precursor, we synthesized either a 40-mer (two cases) or a 38-mer (one case) such that the oligonucleotide was identical to the wild-type sequence of pART2, except at the six positions for the two codons, where an equimolar mixture of all four nucleotide precursors was incorporated during synthesis.

Mutant library construction. The mutagenic oligonucleotides were used to create libraries of random point mutants by using a uracil-containing, single-stranded DNA form of either pART1 or pART2 following the procedure of Kunkel et al. (27). An equimolar ratio of primer to template was used during second-strand synthesis on the uracil-containing template. The mutant clones were recovered as plasmids by transformation of *E. coli* JM101 and selection for ampicillin resistance. The resulting libraries had a mutant frequency of greater than 50%.

Genotype screening. Individual clones were used to inoculate two 1.5-ml cultures to determine their phenotype and genotype. One inoculated culture of supplemented 2XYT (1.0% tryptone, 1.0% yeast extract, 0.5% NaCl, 0.001% thiamine, 100 μg of ampicillin per ml) was grown at 37°C for 4 h. Then 2×10^9 PFU of M13K07 (42) was added, the culture was grown for an additional 1 h, and kanamycin was added to a final concentration of 70 μg/ml. The cultures were grown for an additional 12 to 18 h before harvesting and preparation of single-stranded DNA for DNA sequence analysis as described by Bankier et al. (1).

Phenotype screening. An inoculated culture of supplemented M9 medium (complete M9 plus 0.2% Casamino Acids [Difco Laboratories], 0.001% thiamine, 100 μg of ampicillin per ml) was grown 12 to 18 h prior to induction of the *lac* promoter-operator with 5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After a 1- to 2-h induction, bacteria were lysed in 1/10 volume of loading buffer (62.5 mM Tris hydrochloride [pH 6.8], 2.3% sodium dodecyl sulfate, 10% glycerol, 5% β-mercaptoethanol, 0.05% bromophenol blue) and heated to 100°C for 3 to 4 min. A 10-μl portion of each sample was analyzed on a Western immunoblot as described by Farmerie et al. (8).

RESULTS

Expression of processed *pol* gene products in *E. coli*. We reported previously that expression of a portion of the HIV-1 *pol* gene in bacteria resulted in the appearance of the processed forms of the viral RT (p64 and p51 [8]). These viral proteins were detected by Western blot analysis with serum from an HIV-1-seropositive person. Our original *pol* expression plasmid was truncated in the IN domain, which precluded expression of the IN. To determine whether it was possible to express all of the processed *pol* gene products (PR, p11; the two forms of the RT, p64 and p51; and IN, p34) in bacteria, we expressed all but the first four codons of the *pol* reading frame under the control of an inducible promoter. When an extract prepared from bacteria expressing this clone (pART1) was examined with the HIV-1-positive antiserum, a 34-kDa species was present which comigrated with the presumed viral integrase p34, in addition to the two forms of the viral RT, p64 and p51 (Fig. 1B; compare lane V with lane 1). The p34 species was absent from an extract prepared from cells expressing the clone truncated in the IN

domain (Fig. 1B, lane 2). The RNase H protein was not detected when this antibody reagent was used. When these extracts were examined with a rabbit polyclonal antibody raised against a peptide that includes the carboxyl terminus of the PR and the amino terminus of the RT (2), an 11-kDa protein was seen which comigrated with a viral protein approximately 11 kDa in size (Fig. 1B; compare lanes V, 5, and 6). From this experiment we conclude that the four mature *pol* gene products are made in bacterial cells expressing the HIV-1 *pol* reading frame.

In the bacteria that are expressing p64, p51, p34, and p11 there are several additional discrete bands, which react in the Western blot analysis, the exact sizes of which vary depending on the antibody used. Several of these bands represent reactivity with *E. coli* proteins (Fig. 1B, lane 4 for reactivity with HIV-1-positive human antiserum and lane 9 for reactivity with rabbit antiserum). The rest of the bands appear to be the result of internal initiation at methionine codons. This interpretation is supported by the observation that the intensity of several bands increases when a termination codon or frameshift mutation is introduced near the 5′ end of the *pol* reading frame (Fig. 1B, lane 3 for HIV-1-positive human antiserum and lane 8 for the rabbit antiserum).

In the experiments described below, the appearance of p64, p51, p34, and p11 has been used as an indication of normal processing. The identity of the two forms of the RT has been confirmed by reaction with two separate monoclonal antibodies to the HIV-1 RT (data not shown). The identities of p34 and p11 were confirmed during mutational analysis of the processing sites at their respective amino termini (see below); p34 was also identified by deletion analysis (Fig. 1B, lane 2) and by reactivity with a monospecific rabbit antibody directed against a peptide synthesized to correspond to a portion of the IN domain (data not shown).

Deletion analysis of the *pol* reading frame. In the HIV-1 genome, the 3′ end of the *gag* reading frame overlaps the 5′ end of the *pol* reading frame by 205 or 241 base pairs (205 base pairs for the provirus from which pART1 was derived [36]). The frame shift from the *gag* reading frame to the *pol* reading frame has been shown to occur at codon 3 of the *pol* reading frame (18). The amino terminus of the virion protein believed to be the mature PR is encoded by codon 57 of the *pol* reading frame (30). Ninety-nine codons downstream is the codon for the amino terminus of the mature RT (p64 and p51 [6, 25]). To determine the boundary of the domain required to encode PR function, we constructed a series of deletions at the the 5′ end of the *pol* gene. The two smallest deletions, extending to codons 31 and 40 of the *pol* reading frame, had no effect on the appearance of p64, p51, and p11, the wild-type products that would be detected with the antiserum used in this experiment (Fig. 2, lanes 1 and 2). Deletions beyond this point and up to codon 53 of the *pol* reading frame (lanes 3 to 5) showed the processed *pol* gene products with the concomitant appearance of several larger species. Deletions starting at and beyond codon 56 (lanes 6 to 16) failed to express p11, the mature form of the PR, which was replaced with a series of larger proteins that presumably are the result of degradation of the *pol* precursor by bacterial proteases. With the exception of a deletion extending to codon 59 (lane 8), these deletion mutants also failed to express p64 and p51, and, instead, several new high-molecular-weight species appeared. Thus, deletions to codon 53 do not affect the ability to encode accurate catalytic activity of the PR, nor do they result in the loss of amino acids required for recognition of the N-terminal processing site of the PR.

Missense mutations within a highly conserved domain of the PR-coding region. Toh et al. (40) reported sequence identity between the amino acid sequence of retroviral PRs and the active site of aspartic proteinases. This region of identity is embedded within a larger domain (amino acids 23 through 33) that is conserved among retroviral PRs (Fig. 3). We used oligonucleotide-directed mutagenesis to determine the sensitivity of the HIV-1 PR activity to missense mutations within this relatively conserved domain that includes sequence identity with aspartic proteinases. Mutations were introduced by using a randomly mutagenized oligonucleotide (39 nucleotides in length spanning codons 21 through 33) to prime DNA synthesis on a single-stranded DNA template containing the HIV-1 *pol* gene. We observed three different phenotypes of processing among the mutants, similar to the phenotypes seen in the deletion analysis. First was a phenotype that was indistinguishable from the wild type. Second was a phenotype which showed some appropriately processed products and also some larger unprocessed and partially processed products. Third was a phenotype which showed no normal processing. Examples of the mutant phenotypes are shown in Fig. 4, in which paired conservative and nonconservative substitutions are presented for each amino acid position in this domain.

The phenotypes of all of the mutants tested for this conserved region are shown in Table 1 and can be summarized as follows. Of the 13 codons covered by the mutagenic oligonucleotide, 11 were altered to give 36 independent substitutions; in nine of the positions both conservative and nonconservative substitutions were tested. Of the 36 missense mutations, 25 (69%) resulted in no normal processing. Of the 11 mutations which allowed some level of normal processing, the extent of processing after mutagenesis was similar to the wild type at only two positions (Glu-21 to Val and Asp-30 to Glu). The sequence Asp-Thr-Gly-Ala-Asp (Table 1, positions 25 through 29), which includes the residues with sequence identity to the active site of aspartic proteinases, was extremely sensitive to mutation. All mutations except for one resulted in no normal processing (no missense mutations were recovered at the Gly-27 residue). Only an Ala-to-Ser mutation at position 28 was tolerated, and even this mutant had a very low level of normal processing (Fig. 4, lane 15). At the remaining five positions tested (Ala-22, Leu-23, Leu-24, Thr-31, and Val-32), only conservative changes were tolerated to permit even partial processing (Table 1).

Mutational analysis within the adjacent 13-amino-acid domain (Fig. 3, positions 34 through 46), which is less conserved evolutionarily, showed a lower percentage of aberrant phenotypes (Fig. 5; Table 2). Only 5 of 27 substitutions (19%) resulted in no normal processing. Of the 22 substitutions which permitted processing, 19 resulted in processing that was very similar to that of the wild type. Within this region, only three amino acids are identical between HIV-1 and HIV-2 (Leu-38, Pro-44, and Lys-45), and even these positions could be substituted without affecting processing. These results show that the region of the PR that shares sequence identity with the active site of aspartic proteinases is highly sensitive to mutation, consistent with the hypothesis that HIV-1 PR is an aspartic proteinase.

Mutagenesis of three processing sites within the HIV-1 *pol* precursor protein. At least four processing events take place within the *pol* precursor protein: a cleavage upstream of the

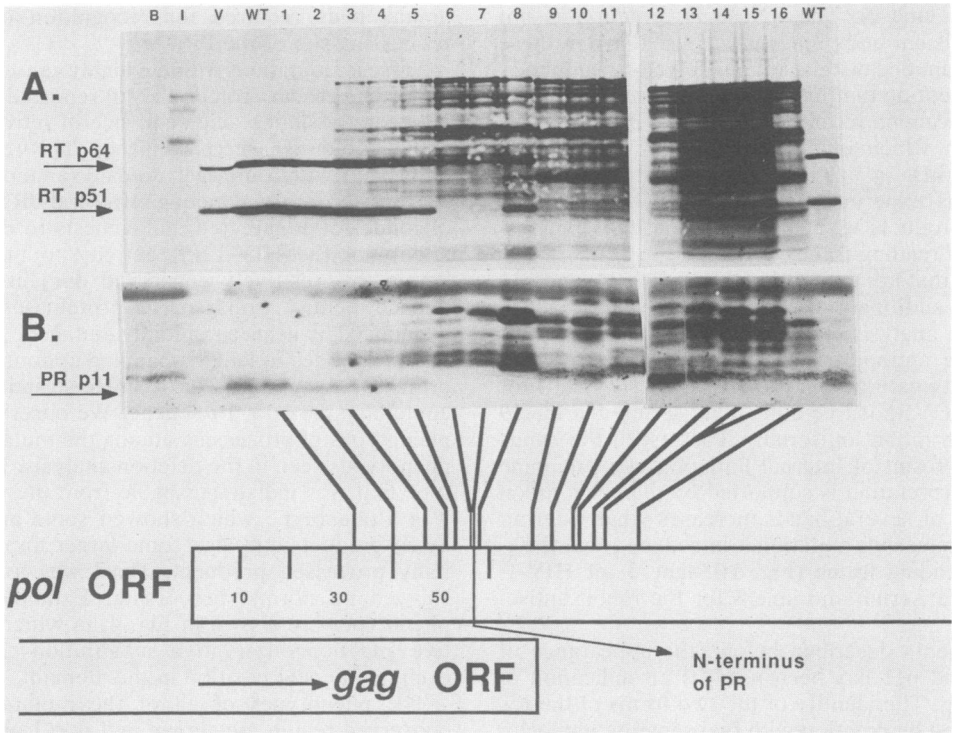

FIG. 2. Deletion analysis of the 5′ end of the *pol* reading frame. The rectangular boxes indicate the arrangement of the *gag* and *pol* open reading frames (ORF), with the position of the N terminus of the virion form of the PR indicated (see text). The numbering inside the *pol* open reading frame indicates the codon number from the beginning of the reading frame. Vertical lines between the top of the box representing the *pol* open reading frame and the lanes in the Western blot identify the deletion endpoints with their corresponding phenotype. Lanes: B, lysate of bacteria containing pIBI20, I, lysate of bacteria expressing pART1E2; V, lysate of pelleted HIV-1; WT, lysate of bacteria expressing pART1; 1 through 16, deletions of the HIV-1 *pol* open reading frame whose right endpoints are indicated by the lines pointing to the box of the *pol* open reading frame below. (A) Western blot of proteins electrophoresed in a 10% polyacrylamide gel, transferred to nitrocellulose paper, and stained with a 1:100 dilution of a monoclonal antibody that recognizes RT p64 and RT p51 (6). (B) Western blot of proteins electrophoresed in a 15% polyacrylamide gel, transferred to nitrocellulose paper, and then stained with a 1:1,000 dilution of a rabbit polyclonal antiserum that recognizes PR p11 (2).

PR to generate the amino terminus of the mature protease; a cleavage between the PR and the RT; a cleavage within the RT to generate its smaller form (p51); and a cleavage between the RT and IN. The positions of three of these cleavages have been identified by amino-terminal sequence analysis of the mature viral protein product (PR, RT, and IN [6, 29, 30]). We have used oligonucleotide-directed mutagenesis to create missense substitutions in the amino acids

```
                1       *10   *    20            ******* ***      40  *
HIV 1           PQITLWQRPLVTIKIGGQLK            EALLDTGADDTVLEEMSLPGRWKPKM
HIV 2           PQfSLWkRPVVTayIeGQpv            EvLLDTGADDSIVagIeLGnnYsPKI
Visna           yvVTe apPkIeIKVGTrwK            klLVDTGADkTIVtshdMSG  iPKgri   i
RSV         lamtMehkDRPLVrViLTntgshpvkqrsvyitALLDSGADiTIIsEedw PtdwPvMeaanpq
BLV         lsipLA rsRPsVAVyLSGpwlqpsqnqa   lmLVDTGAENTVL pqnwlvRdyPRIpaa
MuLV    tlddqGgqGqeppPeprItLkvggqpv         tfLVDTGAQhSVLtQnPgPlsdKSaw

            * * **       60  *           70  *       *     * *** *  *   * 99
HIV 1   IGGIGGFIKVRQYDQILIEIC       GHKAIGTVL VGPTPV   NIIGRNLLTQIGCTLNF
HIV 2   VGGIGGFInKEYkNVeIEVl        nKKvrATIM tGdTPI   NIfGRNILTaLGmSLNl
Visna   LqGIGGiIegekWEQVhLQyk       dKmikGTIVvLATSPV   EVLGRDnMrELGiGLima
RSV     IhGIGGgIpMR ksrdMIELgvi  nrdgslerPLLlfPAvAMvrgsILGRDcLqgLGlrLtnl
BLV     VlGaGGvsRnR YNw  LQgpltlalkPegPfiTIpkIlvdTfdkwQILGRDVLSrLqaSIsip
MuLV    VqGaTGgkRyRwttDrkVhLa       TgKvthSfLhVPdcPy   pLLGRDLLTkLkaqIhFe
```

FIG. 3. Amino acid sequence alignment of the HIV-1 PR domain with other inferred or known retroviral PR domains. Viruses: HIV-1, from pART2 (same in reference 36 except for position 3); HIV-2, HIV type 2 (10); Visna, visna virus (38); RSV, Rous sarcoma virus (37); BLV, bovine leukemia virus (47); MuLV, murine leukemia virus (46). All sequences are compared with the HIV-1 sequence. Amino acid identity between the HIV-1 sequence and a second sequence is indicated with a bold capital letter in the second sequence. A favored substitution (4) between HIV-1 and a second sequence is indicated as a capital letter in the second sequence. Lowercase letters indicate a lack of either an identity or a favored substitution between the second sequence and HIV-1. For positions where all residues are in capitals (indicating only sequence identity and favored substitutions), an asterisk appears above the HIV-1 sequence. The numbering is for the HIV-1 sequence.

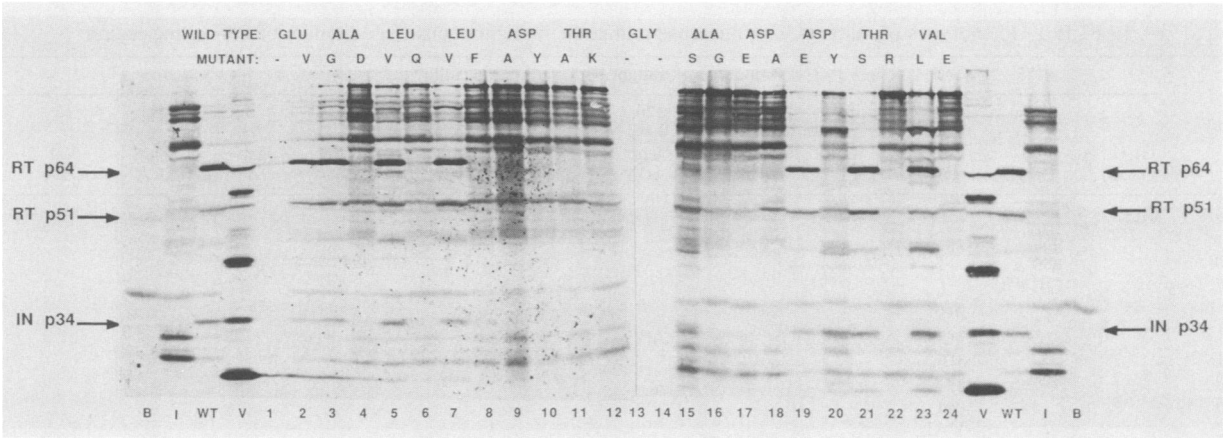

FIG. 4. Missense mutations within a highly conserved domain of the PR-coding region. Lysates of bacteria expressing the pART2 plasmid with the indicated mutations were analyzed by Western blot analysis with a 10% polyacrylamide gel and an HIV-1-positive human antiserum. The wild-type HIV 1 sequence is shown at the top in the three-letter amino acid designation. The mutants tested are shown with the single-letter amino acid designation over the lane which shows the phenotype of that mutant (lanes 1 through 24). Paired conservative and nonconservative changes are shown for each position in the wild-type sequence (residues 21 through 32 [Fig. 3]). A (−) indicates that no mutation was tested at that position. The positions of the two forms of the RT, RT p64 and RT p51, and IN p34 are indicated. Lane B, lysate of bacteria containing pIBI20; lane I, lysate of bacteria expressing pART1E2; lane WT, lysate of bacteria expressing pART2; lane V, lysate of pelleted HIV-1.

flanking the scissile bond (P1 and P1′) to examine sequence requirements of the substrate for cleavage by the viral PR. Examples of the phenotypes of these mutants are shown in Fig. 6, and the phenotypes of all of the mutants tested are summarized in Table 3.

In total, 50 of the possible 400 dipeptide combinations were tested at the three processing sites. Of the 20 amino acids, 17 appeared in P1 (His, Gln, and Tyr were not tested), whereas 18 were tested in P1′ (Val and Glu were not tested). In the substitutions that permitted some level of processing, three amino acids appeared only in P1 (Leu, Val, and Phe), five amino acids appeared only in P1′ (Ser, Thr, Pro, His, and Ile), and two amino acids appeared in both positions (Gly and Ala).

The cleavage site between RT and IN was the most sensitive to mutagenesis. All 17 mutants tested had a single phenotype: a fusion protein between the RT (p64) and IN (p34) (Fig. 6, lanes 7 to 9). There was no apparent effect on the processing of p51. Several minor species were present at the position of p64. The source of these p64-like species is not understood, but could be the result of cleavage at nearby secondary and/or cryptic sites or processing by bacterial enzymes. This background could have obscured a partial cleavage phenotype, although in no case was the IN protein detected.

Nineteen substitutions were tested at the PR-RT cleavage site (examples are shown in Fig. 6, lanes 4 to 6). In two mutants, only one of the two amino acids was changed: Phe-Pro to Phe-Ser permitted partial processing, while Phe-Pro to Glu-Pro blocked processing. Two double substitutions also permitted partial processing. This occurred when either Leu or Val was substituted for Phe in the P1 position and Gly was substituted for Pro in the P1′ position.

The 17 substitution mutants that were tested at the upstream PR cleavage site gave phenotypes that fell into three categories (Fig. 6, lanes 1 to 3). Five of the mutations permitted processing at the upstream cleavage site and at the other cleavage sites to generate RT and IN (Gly-Ala, Gly-His, Leu-Thr, Leu-Pro, and Ala-Ile). The second group of mutants apparently failed to cleave at the mutated upstream PR cleavage site, but cleavage still occurred at the PR-RT and RT-IN cleavage sites (Val-Arg, Arg-Thr, Ser-Thr, Gly-Arg, and Lys-Arg). In these mutants the PR p11 species was absent and was replaced by a series of higher-molecular-weight species (Fig. 6, lane 2). Thus, the processing seen with these mutants was the result of either a small amount (undetected) of mature PR or a larger form of the PR that retained activity. The third group (Leu-Phe, Val-Pro, Asp-Ser, Asn-Ser, Trp-Tyr, Ile-Pro, and Ile-Gln) was similar to the second in that no mature-sized PR was seen; however, with these mutants only a very low level of processing to generate RT p64 and IN occurred (Fig. 6, lane 3).

DISCUSSION

We have used a bacterial expression system to examine the HIV-1 PR encoded within the viral *pol* gene. In this system the HIV-1 *pol* gene is expressed in *E. coli* under the control of an inducible promoter. After induction, four proteins are detected in the bacterial lysate with the same apparent molecular weights as the four processed *pol* gene products found in virions: the PR, p11; two forms of the RT, p64 and p51; and the IN, p34 (Fig. 1). This system provides a simple assay for PR activity, i.e., the generation of processed viral proteins, and permits rapid manipulation of the *pol* gene sequence. We have used this system to probe the sequence requirements for expression of an active PR and PR cleavage site.

The HIV-1 PR domain. We were able to delete the first 52 codons of the *pol* reading frame without destroying PR activity. This is very close to codon 57, which encodes the amino acid found at the amino terminus of the virion protein presumed to be the mature form of the PR (30). Thus, the deletion analysis is consistent with the protein sequence data in defining the HIV-1 PR domain as a small region (99 amino acids based on protein sequence analysis) that lies entirely within the *pol* gene and adjacent to the RT domain. This assignment is further supported by the work of Debouck et al. (5), which showed that a 382-base-pair fragment from this region of the *pol* gene can express a protein in bacteria that has PR activity and is processed to generate the same amino terminus as the protein isolated from the virion.

TABLE 1. Phenotypes of mutants with missing mutations in highly conserved domain of PR-coding region[a]

Mutant	Phenotype of mutant with mutation at following codon[b]:												
	21	22	23 *	24 *	25 *	26 *	27 *	28 *	29 *	30	31 *	32 *	33 *
	GLU GAA	Ala GCT	Leu CTA	Leu TTA	Asp GAT	Thr ACA	Gly GGA	Ala GCA	Asp GAT	Asp GAT	Thr ACA	Val GTA	Leu TTA
M133	**VAL**												
M94		asp											
M170		asp											
M147		GLY											
M86			pro										
M91			arg										
M131			arg										
M150			gln										
M155			gln										
M172			VAL										
M51				MET									
M70				ser									
M101				phe									
M137				phe									
M160				VAL									
M9					tyr								
M84					tyr								
M58					his								
M118					ala								
M121					ala								
M13						ala							
M123						ala							
M46						arg							
M162						ile							
M163						lys							
M35								SER					
M124								SER					
M36								thr					
M61								thr					
M105								thr					
M34								gly					
M132								gly					
M141								glu					
M143								glu					
M184								glu					
M41									val				
M44									ala				
M49									ala				
M76									ala				
M89									ala				
M52									glu				
M145									glu				
M189									glu				
M126									gly				
M149									gly				
M187									asn				
M1										**GLU**			
M144										**GLU**			
M148										**GLU**			
M7										ASN			
M21										TYR			
M127										TYR			
M175										TYR			
M73											ala		
M151											SER		
M178											arg		
M85											ile		
M95											ile		
M166											ile		
M26												LEU	
M83												LEU	
M93												glu	

[a] Amino acid and codon for wild-type strian are shown below codon numbers for positions 21 to 33. Asterisks denote conserved residues (see Fig. 3).

[b] Underlined boldface capitals denote identity with the wild type; lightface capitals denote intermediate level of processing; lowercase letters denote no normal processing.

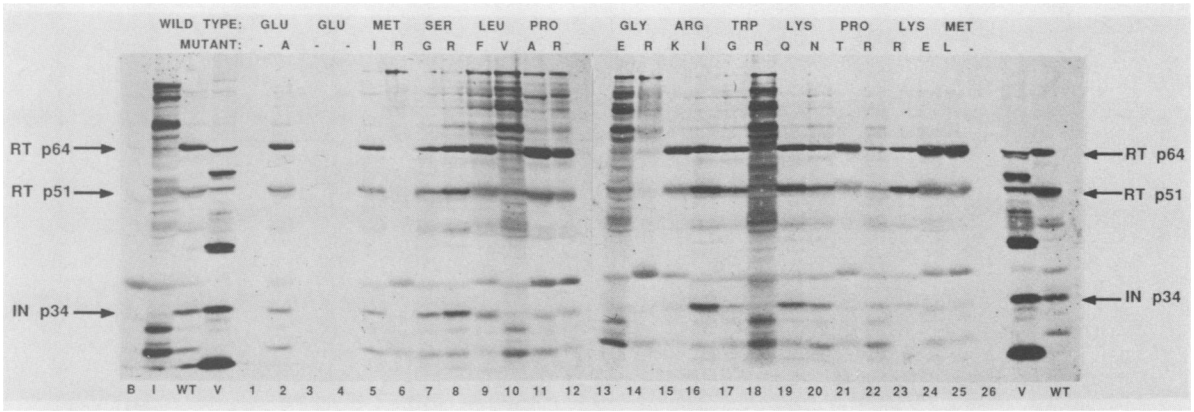

FIG. 5. Missense mutations within a poorly conserved domain of the PR-coding region. Analysis similar to that described in the legend to Fig. 4 was carried out for point mutations in a poorly conserved domain of the protease (residues 34 to 46 [Fig. 3]). All procedures and notations are as described in the legend to Fig. 4.

Significance of sequence identity with aspartic proteinases. Retroviral PRs show sequence identity with the active site of aspartic proteinases (40). This similarity is strengthened by the observation that pepstatin, an inhibitor specific to aspartic proteinases, inhibits several retroviral PRs, although inhibition is apparent only at relatively high concentrations of the inhibitor (11, 21). On the basis of the assumption that the active site of an enzyme should be one of the most sensitive sites to mutation with respect to enzyme activity, we have undertaken extensive mutagenesis of the HIV-1 PR.

TABLE 2. Phenotypes of mutants with missense mutations in less highly conserved domain[a]

Mutant	Phenotype of mutant with mutation at following codon[b]:												
	34	35	36	37	38	39	40	41	42	43	44 *	45	46
	Glu GAA	Glu GAA	Met ATG	Ser AGT	Leu TTG	Pro CCA	Gly GGA	Arg AGA	Trp TGG	Lys AAA	Pro CCA	Lys AAA	Met ATG
Q24	**ALA**												
Q92			**ILE**										
140			arg										
Q22				**ARG**									
Q66				**GLY**									
Q3					PHE								
Q25					VAL								
Q2						ARG							
Q116						**ALA**							
Q128						**THR**							
Q139						**LEU**							
Q42							glu						
Q61							glu						
Q148							arg						
Q18								**ILE**					
Q49								**LYS**					
Q93								**LYS**					
Q110								**LYS**					
Q40									arg				
Q59									arg				
Q76									**GLY**				
Q82									**GLY**				
Q101									leu				
Q106									leu				
Q43										**GLN**			
Q77										**ASN**			
Q109										**ASN**			
Q105										**GLU**			
Q65											**ARG**		
Q108											**THR**		
Q104												**ARG**	
Q107												**GLN**	
Q136												**THR**	
Q147													**LEU**

[a] Amino acid and codon for wild-type strain are shown below codon numbers for positions 34 to 46. For definition of asterisk, see Table 1, footnote *a*.
[b] See Table 1, footnote *b*.

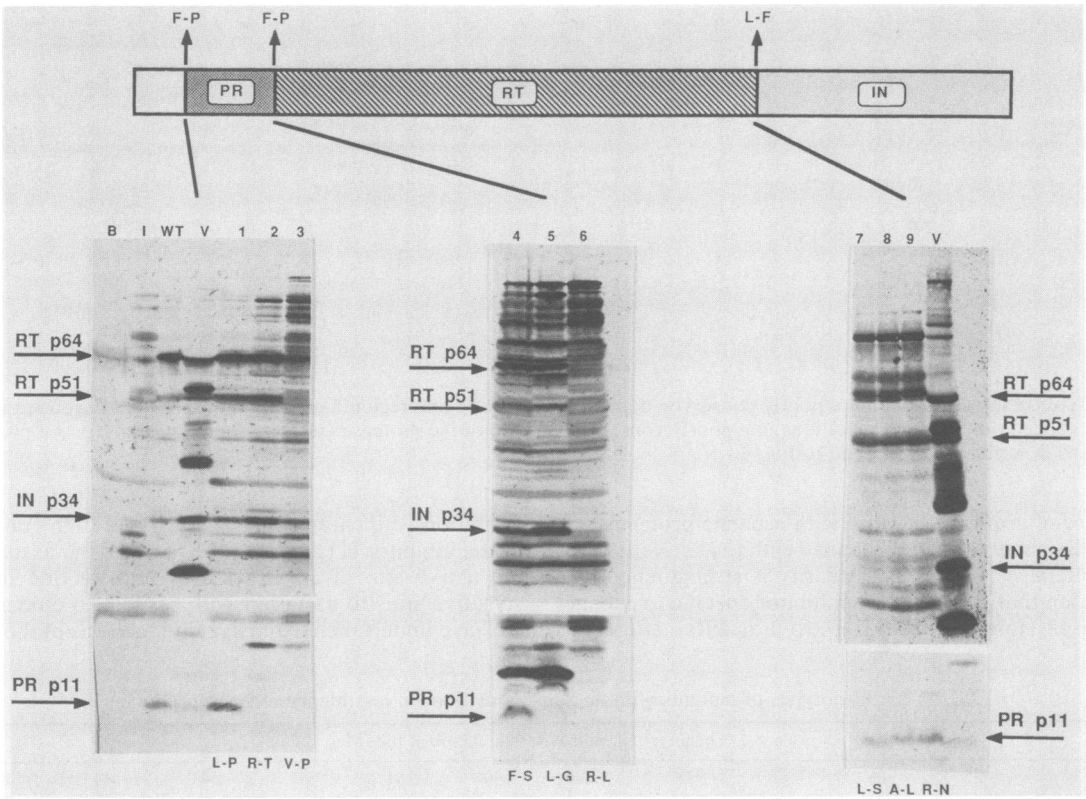

FIG. 6. Mutagenesis of three processing sites within the HIV-1 *pol* precursor protein. The box at the top represents the *pol* reading frame with its four domains: the region upstream of the PR, the PR domain, the RT domain, and the IN domain. The amino acids flanking the scissile bonds that separate each domain are shown above the box. The positions of the processing sites are inferred from protein sequence data (see text). Below each box are Western blot analyses of examples of *pol* gene mutants with altered processing site sequences. The altered sequence is shown at the bottom of the lane for lanes 1 through 9. The upper panel of each blot represents analysis of mutants expressed in bacteria with the bacterial lysate electrophoresed in a 10% polyacrylamide gel and reacted with an HIV-1-positive human antiserum. The lower panel is an analysis of the same lysates in a 15% polyacrylamide gel and reaction with the rabbit antibody to PR-RT (2). Positions of the wild-type products, PR (PR p11), RT (RT p64 and RT p51), and IN (IN p34) are indicated. Lane B, lysate of bacteria containing pIBI20; lane I, lysate of bacteria expressing pART1E2; lane WT, lysate of bacteria expressing pART2; lane V, lysate of pelleted HIV-1.

In our initial screen, we mutagenized two adjacent 13-amino-acid domains (approximately 25% of the PR domain) and examined the activity of 96 mutants. The first 13-amino-acid domain, which spans a highly conserved region among retroviruses and contains the residues that show sequence identity with the active site of aspartic proteinases (Asp-Thr-Gly, residues 25 to 27), was very sensitive to mutagenesis: 69% of the substitution mutants tested failed to carry out normal processing of the *pol* precursor (Fig. 4; Table 1). This high level of sensitivity is contrasted with the lower level of sensitivity to mutagenesis seen in the adjacent 13-amino-acid domain: only 22% of the substitution mutants in this region could not process the *pol* precursor (Fig. 5; Table 2). Thus, the region of the HIV-1 PR that shows sequence identity with the active site of aspartic proteinases is very sensitive to mutagenesis, consistent with the idea that this region of the PR is part of the active site and consistent with the hypothesis that the HIV-1 PR shares a catalytic mechanism with aspartic proteinases. A similar approach has been used in which mutation of Asp-25 was shown to abolish processing by the HIV-1 PR by using either a bacterial expression system (23, 33) or a mutated virus (23), and an Asp-to-Ile mutation was shown to block the activity of the Rous sarcoma virus PR (25).

There are important differences between aspartic proteinases and the HIV-1 PR, most notably in size. Aspartic proteinases consist of two similar lobes that are thought to have arisen by gene duplication (39). Each lobe donates an aspartic acid residue to the active site. The HIV-1 PR is much smaller than the typical aspartic proteinase (11 versus 35 kDa), making it unlikely that one molecule of the viral PR could provide both aspartic acid residues to the active site. This has led to the suggestion that the PR must first dimerize to form an active enzyme (35). The need to dimerize may provide the virus with a mechanism for controlling the activity of its PR. In at least one case it has been shown that processing of viral proteins does not occur until the time of virion formation (45). Such control of processing could be achieved if the concentration of the *gag-pol* precursor is not sufficiently high within the cell to permit dimer formation until incorporation within the virion. When the *gag-pol* precursor is concentrated during virion formation, dimerization of the PR domain could occur, and processing of viral proteins would proceed.

The PR cleavage site. The cleavage sites for retroviral PRs have been deduced largely from the amino-terminal sequence of viral *gag* and *pol* proteins and by comparison of the equivalent positions in genomes for which only the nucleotide sequence is available (41). The cleavage sites represent a family of sequences that are generally hydrophobic. Several schemes have been suggested to describe amino acids at specific positions around the scissile bond. Oroszlan

TABLE 3. Phenotypes of processing site mutants within HIV-1 *pol* precursor protein

Upstream PR (Phe-Pro)			PR-RT (Phe-Pro)			RT-IN (Leu-Phe)		
Mutant	Substitution	Processing Level	Mutant	Substitution	Processing Level	Mutant	Substitution	Processing Level
T4	Gly-Ala	Wild type	R5	Val-Gly	Low	S1	Trp-Leu	Negative
T12	Gly-His	Wild type	R17	Leu-Gly	Low	S3	Ile-Ser	Negative
T19	Leu-Thr	Wild type	R9	Phe-Ser	Low	S6	Ile-Ala	Negative
T28	Leu-Pro	Wild type	R2	Trp-Gly	Negative	S8	Thr-Gly	Negative
T30	Ala-Ile	Wild type	R3	Arg-Leu	Negative	S11	Ala-Asp	Negative
T8	Val-Arg	Efficient[a]	R7	Leu-Thr	Negative	S12	Phe-Thr	Negative
T10	Arg-Thr	Efficient[a]	R10	Gly-Ser	Negative	S13	Gly-Arg	Negative
T23	Ser-Thr	Efficient[a]	R11	Glu-Pro	Negative	S14	Arg-Leu	Negative
T25	Gly-Arg	Efficient[a]	R13	Ser-His	Negative	S15	Ser-Met	Negative
T26	Lys-Arg	Efficient[a]	R15	Leu-Leu	Negative	S17	Leu-Ser	Negative
T3	Leu-Phe	Low[a]	R16	Ser-Met	Negative	S18	Met-Thr	Negative
T5	Val-Pro	Low[a]	R18	Thr-Leu	Negative	S20	Glu-Lys	Negative
T11	Asp-Ser	Low[a]	R21	Glu-Tyr	Negative	S22	Ser-Ser	Negative
T14	Asn-Ser	Low[a]	R22	Leu-Arg	Negative	S26	Cys-Met	Negative
T15	Trp-Tyr	Low[a]	R23	Ala-Lys	Negative	S28	Arg-Asn	Negative
T18	Ile-Pro	Low[a]	R25	Gly-Cys	Negative	S29	Ala-His	Negative
T20	Ile-Gln	Low[a]	R26	Cys-Lys	Negative	S30	Ala-Leu	Negative
			R27	Gly-Ile	Negative			
			R29	Pro-Trp	Negative			

[a] No upstream cleavage.

and Copeland (34) have proposed the following features of the cleavage site: P4 (the fourth amino acid upstream of the scissile bond) is the beta turner; P3 is polar; and P2 is hydrophobic. Pearl and Taylor (L. H. Pearl and W. R. Taylor, Letter, Nature (London) **328**:482, 1987) have suggested another scheme for those sites in which Pro is in the P1′ site: P2 is small; P1 is aromatic or large and hydrophobic; and P2′ (the second amino acid downstream of the scissile bond) is small and hydrophobic. A more specific sequence pattern has recently been suggested for primate lentivirus cleavage sites (14). Deletion analysis of the HIV-1 *pol* reading frame upstream of the PR cleavage site showed that as few as four virus-specific amino acids are needed to define the cleavage site to the left of the scissile bond (Fig. 2). In an initial attempt to define the sequence requirement of the cleavage site, we have also mutagenized the P1 site and the P1′ site of three PR cleavage sites in the HIV-1 *pol* precursor (Fig. 6; Table 3). Two of these sites (upstream PR and PR-RT) have Phe-Pro as the normal sequence, whereas the third site (RT-IN) is Leu-Phe.

Although the data set for mutations at the three cleavage sites is complex and incomplete, several points are apparent. The first point is that certain types of amino acids are tolerated in the P1 and the P1′ sites. In P1, Gly and the hydrophobic amino acids Ala, Val, Leu, and Phe are present in the various substitutions that permit processing (although Val appears only in the weakly processed mutant R5). The amino acids Gly, Thr, Pro, Ser, Ala, His, and Ile are present in P1′ in cleavable mutants. With the exception of Ile and His, the amino acids in the P1′ site represent members of a group of amino acid equivalents as defined by examining evolutionarily related proteins (4). Like the other amino acids in this group, Ile and His would be largely uncharged at neutral pH. A more precise definition of the permitted amino acids at these and surrounding positions will be obtained by examination of a larger number of mutants. The second point that can be made from these data is that the three cleavage sites are not equivalent with respect to the P1 and P1′ sites. There are two examples, including one wild-type sequence, in which a dipeptide sequence functioned at one cleavage site but blocked cleavage at another processing site. Leu-Phe, the wild-type sequence at the RT-IN site, blocked cleavage when substituted at the upstream PR site; and Leu-Thr was cleaved when substituted at the upstream PR site but blocked cleavage when substituted at the PR-RT site (Table 3). Thus, the amino acids flanking the scissile bond are seen by the PR in a broader context that must include surrounding amino acids.

Finally, two classes of mutants failed to process at the upstream PR cleavage site (Table 3). One class processed the other sites in the *pol* precursor efficiently, whereas the second class did not. Examination of the amino acid substitutions that make up these two classes reveals that the first class contains positively charged amino acids in four of the five mutants (the fifth is a Ser-Thr substitution), whereas the second class contains pairs of amino acids, of which one has a carbonyl group in the side chain, or pairs of hydrophobic amino acids. One possible interpretation of these results is that the positively charged residues (Arg and Lys) preclude any binding to the PR substrate-binding site, leaving it available to process at other sites. The second group of mutants may mimic the normal cleavage site sufficiently to permit binding by the PR but then remain in the substrate-binding site and block the PR from cleavage at other sites. Kotler et al. (24) have observed previously that the introduction of Ile into the P1 site of a peptide substrate blocked cleavage by the avian myeloblastosis virus PR and that this peptide also functioned as an inhibitor.

The fact that inhibition of cleavage can occur at nonmutated sites when the upstream PR site is mutated but not when the RT-IN site is mutated suggests that there may be ordered cleavage, with the upstream PR site cleaved before the RT-IN site. The extreme phenotype seen in mutants with mutations at the PR-RT site (Fig. 6, lanes 4 to 6) suggests that this site may be cleaved first; this conclusion is consistent with the structure of a reported processing intermediate (11).

ACKNOWLEDGMENTS

We thank Ron Luftig for helpful discussion, and we are especially indebted to Philip Barr and Fulvia Di Marzo Veronese for providing antibody reagents, to Lee Ratner for providing the DNA clone of

HIV-1, and to Robert Buckheit for providing pelleted virus. Virginia White and John O'Bryan provided technical assistance in some of the Western blot analyses, and Bruce Wisely synthesized and purified the oligonucleotides used in the study. The efforts of Elizabeth Craig in preparing the manuscript are gratefully acknowledged.

This work was supported by Public Health Service grants AI23951, AI25321, AI/CA25697, AI08998, and GM21313 from the National Institutes of Health and by an award from the North Carolina United Way. R.S. is supported by a Junior Faculty Award from the American Cancer Society.

LITERATURE CITED

1. **Bankier, A. T., K. M. Weston, and B. G. Barrell.** 1987. Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. Methods Enzymol. **155**:51–93.
2. **Barr, P. J., M. D. Power, C. T. Lee-Ng, H. L. Gibson, and P. A. Luciw.** 1987. Expression of active human immunodeficiency virus reverse transcriptase in *Saccharomyces cerevisiae*. Bio/Technology **5**:486–489.
3. **Crawford, S., and S. P. Goff.** 1985. A deletion mutant in the 5′ part of the *pol* gene of Moloney murine leukemia virus blocks proteolytic processing of the *gag* and *pol* polyproteins. J. Virol. **53**:899–907.
4. **Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt.** 1978. A model of evolutionary change in proteins, p. 345–352. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5. National Biomedical Research Foundation, Washington, D.C.
5. **Debouck, C., J. G. Gorniak, J. E. Strickler, T. D. Meek, B. W. Metcalf, and M. Rosenberg.** 1987. Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the *gag* precursor. Proc. Natl. Acad. Sci. USA **84**:8903–8906.
6. **Di Marzo Veronese, F., T. D. Copeland, A. L. DeVico, R. Rahman, S. Oroszlan, R. C. Gallo, and M. G. Sarngadharan.** 1986. Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. Science **231**:1289–1291.
7. **Dittmar, K. J., and K. Moelling.** 1978. Biochemical properties of p15-associated protease in an avian RNA tumor virus. J. Virol. **28**:106–118.
8. **Farmerie, W. G., D. D. Loeb, N. C. Casavant, C. A. Hutchison III, M. H. Edgell, and R. Swanstrom.** 1987. Expression and processing of the AIDS virus reverse transcriptase in *Escherichia coli*. Science **236**:305–308.
9. **Graves, M. C., J. J. Lim, E. P. Heimer, and R. Kramer.** 1988. An 11-kDa form of human immunodeficiency virus protease expressed in *Escherichia coli* is sufficient for enzymatic activity. Proc. Natl. Acad. Sci. USA **85**:2449–2453.
10. **Guyader, M., M. Emerman, P. Sonigo, F. Clavel, L. Montagnier, and M. Alizon.** 1987. Genome organization and transactivation of the human immunodeficiency virus type 2. Nature (London) **326**:662–669.
11. **Hansen, J., S. Billich, T. Schulze, S. Sukrow, and K. Moelling.** 1988. Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. EMBO J. **7**:1785–1791.
12. **Hansen, J., T. Schulze, W. Mellert, and K. Moelling.** 1988. Identification and characterization of HIV-specific RNase H by monoclonal antibody. EMBO J. **7**:239–243.
13. **Hansen, J., T. Schulze, and K. Moelling** 1987. RNase H activity associated with bacterially expressed reverse transcriptase of human T-cell lymphotropic virus III/lymphadenopathy-associated virus. J. Biol. Chem. **262**:12393–12396.
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. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
16. **Hoheisel, J., and F. M. Pohl.** 1986. Simplified preparation of unidirectional deletion clones. Nucleic Acids Res. **14**:3605.
17. **Hutchison, C. A., III, S. K. Nordeen, K. Vogt, and M. H. Edgell.** 1986. A complete library of point substitution mutations in the glucocorticoid response element of mouse mammary tumor virus. Proc. Natl. Acad. Sci. USA **83**:710–714.
18. **Jacks, T., M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, and H. E. Varmus.** 1988. Characterization of ribosomal frameshifting in HIV-1 *gag/pol* expression. Nature (London) **331**:280–283.
19. **Jacks, T., K. Townsley, H. E. Varmus, and J. Majors.** 1987. Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus *gag*-related polyproteins. Proc. Natl. Acad. Sci. USA **84**:4298–4302.
20. **Jacks, T., and H. E. Varmus.** 1985. Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science **230**:1237–1242.
21. **Katoh, I., T. Yasunaga, Y. Ikawa, and Y. Yoshinaka.** 1987. Inhibition of retroviral protease activity by an aspartyl proteinase inhibitor. Nature (London) **329**:654–656.
22. **Katoh, I., Y. Yoshinaka, A. Rein, M. Shibuya, T. Odaka, and S. Oroszlan.** 1985. Murine leukemia virus maturation: protease region required for conversion from "immature" to "mature" core form and for virus infectivity. Virology **145**:280–292.
23. **Kohl, N. E., E. A. Emini, W. A. 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.
24. **Kotler, M., R. A. Katz, W. Dahno, J. Leis, and A. M. Skalka.** 1988. Synthetic peptides as substrates and inhibitors of a retroviral protease. Proc. Natl. Acad. Sci. USA **85**:4185–4189.
25. **Kotler, M., R. A. Katz, and A. M. Skalka.** 1988. Activity of avian retroviral protease expressed in *Escherichia coli*. J. Virol. **62**:2696–2700.
26. **Kramer, R. A., M. D. Schaber, A. M. Skalka, K. Ganguly, F. Wong-Staal, and E. P. Reddy.** 1986. HTLV-III *gag* protein is processed in yeast cells by the virus *pol*-protease. Science **231**:1580–1584.
27. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154**:367–382.
28. **Le Grice, S. F. J., V. Beuck, and J. Mous.** 1987. Expression of biologically active human T-cell lymphotropic virus type III reverse transcriptase in *Bacillus subtilis*. Gene **55**:95–103.
29. **Lightfoote, M. M., J. E. Coligan, T. M. Folks, A. S. Fauci, M. A. Martin, and S. Venkatesan.** 1986. Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus. J. Virol. **60**:771–775.
30. **Lillehoj, E. P., R. H. R. Salazar, R. J. Mervis, M. G. Raum. H. W. Chan, N. Ahmad, and S. Venkatesan.** 1988. Purification and characterization of the putative *gag/pol* protease of human immunodeficiency virus. J. Virol. **62**:3053–3058.
31. **Mermer, B., M. Malamy, and J. M. Coffin.** 1983. Rous sarcoma virus contains sequences which permit expression of the *gag* gene in *Escherichia coli*. Mol. Cell. Biol. **3**:1746–1758.
32. **Moore, R., M. Dixon, R. Smith, G. Peters, and C. Dickson.** 1987. Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of *gag* and *pol*. J. Virol. **61**:480–490.
33. **Mous, J., E. P. Heimer, and S. F. J. LeGrice.** 1988. Processing protease and reverse transcriptase from human immunodeficiency virus type 1 polyprotein in *Escherichia coli*. J. Virol. **62**:1433–1436.
34. **Oroszlan, S., and T. D. Copeland.** 1985. Primary structure and processing of *gag* and *env* gene products of human T-cell leukemia viruses $HTLV\text{-}I_{CR}$ and $HTLV\text{-}I_{ATK}$. Curr. Top. Microbiol. Immunol. **115**:221–233.
35. **Pearl, L. H., and W. R. Taylor.** 1987. A structural model for the retroviral proteases. Nature (London) **329**:351–354.
36. **Ratner, L., A. Fisher, L. L. Jagodzinski, H. Mitsuya, R. S. Liou, R. C. Gallo, and F. Wong-Staal.** 1987. Complete nucleotide sequences of functional clones of the AIDS virus. AIDS Res. Hum. Retroviruses **3**:57–69.
37. **Sauer, R. T., D. W. Allen, and H. D. Niall.** 1981. Amino acid

sequence of p15 from avian myeloblastosis virus complex. Biochemistry **20**:3784–3791.

38. **Sonigo, P., M. Alizon, K. Staskus, D. Klatzmann, S. Cole, O. Danos, E. Retzel, P. Tiollais, A. Hease, and S. Wain-Hobson.** 1985. Nucleotide sequence of the Visna lentivirus: relationship to the AIDS virus. Cell **42**:369–382.
39. **Tang, J., M. N. G. James, I. N. Hsu, J. A. Jenkins, and T. L. Blundell.** 1978. Structural evidence for gene duplication in the evolution of the acid proteases. Nature (London) **271**:618–621.
40. **Toh, H., R. Kikuno, H. Hayashida, T. Miyata, W. Kugimiya, S. Inouye, S. Yuki, and K. Saigo.** 1985. Close structural resemblance between putative polymerase of a Drosophila transposable genetic element 17.6 and *pol* gene product of Moloney murine leukaemia virus. EMBO J. **4**:1267–1272.
41. **Van Beveren, C., J. Coffin, and S. Hughes.** 1985. Appendixes, p. 559–1221. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), Molecular biology of tumor viruses: RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
42. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153**:3–11.
43. **Vogt, V. M., W. Wight, and R. Eisenman.** 1979. In vitro cleavage of avian retrovirus *gag* proteins by viral protease p15. Virology **98**:154–167.
44. **Von der Helm, K.** 1977. Cleavage of Rous sarcoma viral polyprotein precursor into internal structural proteins in vitro involves viral protein p15. Proc. Natl. Acad. Sci. USA **74**:911–915.
45. **Witte, O., and D. Baltimore.** 1978. Relationship of retrovirus polyprotein cleavage to virion maturation studied with temperature-sensitive murine leukemia virus mutants. J. Virol. **26**:750–761.
46. **Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan.** 1985. Murine leukemia virus protease is encoded by the *gag/pol* gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA **82**:1618–1622.
47. **Yoshinaka, Y., I. Katoh, T. D. Copeland, G. W. Smythers, and S. Oroszlan.** 1986. Bovine leukemia virus protease: purification, chemical analysis, and in vitro processing of *gag* precursor polyproteins. J. Virol. **57**:826–832.
48. **Yoshinaka, Y., and R. B. Luftig.** 1977. Properties of a P70 proteolytic factor of murine leukemia viruses. Cell **12**:709–719.