The protease of human immunodeficiency virus has been expressed in *Escherichia coli* and purified to apparent homogeneity. Immunoreactivity toward anti-protease peptide sera copurified with an activity that cleaved the structural polyprotein gag p55 and the peptide corresponding to the sequence gag 128–135. The enzyme expressed as a nonfusion protein exhibits proteolytic activity with a pH optimum of 5.5 and is inhibited by the aspartic protease inhibitor pepstatin with a $K_i$ of 1.1 $\mu$M. Replacement of the conserved residue Asp-25 with an Asn residue eliminates proteolytic activity. Analysis of the minimal peptide substrate size indicates that 7 amino acids are required for efficient peptide cleavage. Size exclusion chromatography is consistent with a dimeric enzyme and circular dichroism spectra of the purified enzyme are consistent with a proposed structure of the protease (Pearl, L. H., and Taylor, W. R. (1987) *Nature* 329, 351–354). These data support the classification of the human immunodeficiency virus protease as an aspartic protease, likely to be structurally homologous with the well characterized family that includes pepsin and renin.

The etiological agent of acquired immune deficiency syndrome (AIDS) is the retrovirus human immunodeficiency virus (HIV). HIV proteins, like other retroviral proteins, are initially translated as the large precursor polypeptides gag, pol, and env and are proteolytically processed to generate structural proteins (p17, p24, p7, and p6), enzymes (protease, reverse transcriptase, and integrase), and the envelope proteins (gp120 and gp41). Processing of the gag and pol polyproteins is believed to involve a virally encoded protease. Among the retroviral proteases that have been characterized are those from the avian myeloblastosis virus, bovine leukemia virus, and murine leukemia virus (2–5). Mutations within the protease coding regions of retroviruses have been shown to result in noninfectious virions (6), including the case of HIV-1 (7).

There is limited amino acid sequence similarity between the well characterized aspartic proteases (such as pepsin and renin) and the HIV protease. Based on this similarity which involves the peptide sequences that flank the critical aspartate residues of the pepsins (8), the retroviral proteases have been speculated to belong to the family of aspartic proteases. Whereas retroviral proteases range in size from 10 to 14 kDa, the aspartic proteases are commonly observed to have masses of 33–44 kDa. Observation of similarity in the amino acid sequence between the 2 domains of the pepsin family of proteases suggested they have evolved from a duplication of genes (9). The smaller retroviral proteases have thus been postulated to function as dimers in analogy to the 2-domain structure of the pepsin class (10, 11).

The proteolytic activity of the HIV-1 protease has been observed with *in vivo* self-processing of portions of the pol gene product following expression in microbial cells and intermolecular reactions performed with that product have been observed with microbially expressed gag p55 as substrate (12–15). Attempts have been made to purify the microbially expressed protease, and partially purified material was used to demonstrate inhibition of the gag p55 cleavage reaction by 1 mM pepstatin (15). In addition, while this paper was in preparation, the purification of the protease from virions was reported, although the purified material had no activity (16). In this report, we describe the expression of the protease as a 10-kDa nonfusion protein in *Escherichia coli* and its purification as a soluble, active enzyme to apparent homogeneity. The activity of the purified protease with peptide substrates and its inhibition by pepstatin is quantitatively described. Furthermore, we present preliminary characterization of the protease structure with gel filtration and circular dichroism.

**EXPERIMENTAL PROCEDURES**

**Materials**—Peptides used as substrates or in the generation of antisera were synthesized by the Merrifield solid phase synthesis method (17). The substrate protein, HIV-1 gag p55, was cloned and expressed in yeast. Chemically synthesized protease was kindly donated by Drs. R. Nutt and D. Veber (Merck Sharp and Dohme). The mono S column and plasmid pKK233-2 were from Pharmacia LKB Biotechnology Inc., HIV-1 viral lysate from Cytotech (San Diego, CA), and pepstatin A from Sigma. Other chemicals were reagent grade from standard suppliers. Rabbit antisera were raised against the peptide corresponding to pol open reading frame residues 102–114 (protease residues 34–46) by established methods (18). Oligonucleotides were synthesized using an Applied Biosystems model 380A DNA synthesizer.

**Cloning and Expression**—Recombinant DNA procedures were performed as described by Maniatis et al. (19). The 5’ portion (HindIII to KpnI) of the pol gene of the NY5 (20) strain of HIV-1 was subcloned into the HindIII/KpnI site of M13 mp18 and mp8. A mutation was introduced by oligonucleotide-directed mutagenesis to create an NcoI site.
site at codon positions 68 and 69 of the open reading frame. Briefly, the oligonucleotide (5'-GATCCTTTACATAGCTTCAG) was annealed to the single stranded DNA as a primer, extended with Klenow DNA polymerase and T4 DNA ligase, and digested with BglII/KpnI. The mutated fragment was subcloned into BamHI/KpnI-digested pUC19 and transformed into E. coli strain DH5α. Ampicillin-resistant colonies were screened by colony hybridization using the mutant oligonucleotide as a probe. Plasmid DNA was isolated from hybridization-positive clones and the mutation resolved by subsequent retransformation. The Ncol/AhaIII fragment from the mutant plasmid encoding pol amino acids 68-165 was mixed with a pair of complementary oligonucleotides (5'-GAATTCTTACA and 5'-ATGTCGAGAAAGTT) and with the Ncol/HindIII-digested pGEM4 which had been modified by insertion of an Ncol linker into the Smal site of the polynucleotid. The mixture was ligated and transformed into E. coli strain DH5α. Ampicillin-resistant colonies were screened by hybridization with probes for both the added linker and the protease fragment. The Ncol/HindIII fragment from the resultant plasmid which contained the protease gene was cloned into the Ncol/HindIII site of pTRP and transformed into E. coli strain DH5α to yield the plasmid pPRT. pPRT is a derivative of pKK233-2 in which the EcoRI/Ncol fragment containing the tac promoter had been replaced by a synthetic EcoRI/Ncol fragment containing the trp promoter and a double cistrionic ribosome binding site (Fig. 1). pPRT encodes amino acids 2-99 of the protease preceded by Met-Ala residues.

Construction of the Met-Ala sequence described above was convenient due to the introduction of the Ncol site. However, the viral genome encodes a Pro at the suspected NH2-terminus (pol open reading frame codon 68) of the protease, so site-directed mutagenesis was employed to change Ala-1 to Pro. The Ncol/HindIII fragment for pPRT was isolated, blunted with Klenow DNA polymerase, and BamHI linkers were added to the fragment. The linker DNA was cloned into the BamHI site of pUC13 and the BamHI fragment isolated and partially digested with DdeI, which cuts after the HindIII site (Fig. 1). pPRT encodes amino acids 2-99 of the protease preceded by Met-Ala residues.

Fig. 1. Map of the protease expression plasmid pPRT. The location of the protease is indicated beginning with the ATG and ending with TAA. PTrp is the trp promoter. The large circle between the EcoRI and HindIII sites is derived from the plasmid pKK233-2 and includes the E. coli origin of DNA replication, the β-lactamase, and the transcriptional terminator (S/JTR17).
used for Gln-Asn-Tyr-Pro-Ile-Val. The program Enzfit was used to fit data to simple Michaelis-Menten kinetics (23).

Protease activity as a function of pH was measured using 10 mM substrate (Val-Ser-Gln-Asn-Tyr-Pro-Ile-Val) in the following 50 mM buffers: pH 3.0, 3.5, 4.0, sodium formate; pH 4.0, 4.5, 5.0, 5.5, NaOAc; pH 5.5, 6.0, 6.5, NaMES; pH 6.5, 7.0, NaMOPS. A 6-µl aliquot from a 4.1 µg/ml solution of the purified [α-Ala-1] protease was employed.

**Electrophoresis and Immunoblots**—SDS-polyacrylamide gel electrophoresis was run according to Laemmli (22) using 16% gels and used for to fit data to simple Michaelis-Menten kinetics (23).

**Gel Filtration**—Sephadex G-75, 40–120 µm, in a 2.2 × 100-cm column was used to estimate Stokes radii with the following calibration standards: lysozyme (14.4 kDa), cytochrome c (12.7 kDa), α-chymotrypsinogen (25 kDa), carboxylic anhydride (30 kDa), and ovalbumin (43 kDa).

**Circular Dichroism Spectra**—A solution containing HIV protease (0.8 mg, 1 ml) was dialyzed overnight against 2 liters of 50 mM NaMES, pH 5.5, 0.1% DTT, 0.25 mM EDTA, and 10% glycerol. Circular dichroism measurements were conducted using a Jasco model J-41A spectropolarimeter over the wavelength range of 254–194 nm using 0.1 and 0.02-cm path length cells at 25 °C. The data were electronically captured and base-line subtraction and conversion to mean residue ellipticity versus wavelength were calculated by computer. Estimation of the relative amounts of secondary structure was performed using the program from the Chang-Wu-Yang method (26) described by Yang et al. (27).

**RESULTS**

To express the HIV-1 protease in *E. coli*, the gene fragment that encoded the putative protease was modified so as to encode within an *NcoI* to HindIII restriction fragment a polypeptide that initiates at the NH₂ terminus with a Met-Val-Ala followed by residues 70–167 of the pol open reading frame terminating with a stop codon. This gene fragment was introduced behind the trp promoter contained within a pBR322-derived vector, resulting in the HIV-1 protease expression vector pPRT. Fortuitously, during the construction of the desired expression plasmid pPRT, we isolated a similar DNA construct that lacked the codon for the initial ATG (Met) codon (pATG-).

The plasmid pPRT was used to transform HB101 to ampicillin resistance. Expression of the HIV-1 protease was then induced by growing the transformed strain in tryptophan-depleted medium in the presence of indoleacrylic acid. Unlike the control strain transformed with either the vector alone or the vector pATG-, the strain containing pPRT was inhibited in its growth under conditions that induced the trp promoter. When cell lysates were analyzed by immunoblotting with antisera raised against a peptide corresponding to amino acids 102–114 of the pol open reading frame, an immunoreactive band of 10 kDa was observed in the pPRT sample but not in control strains transformed with pATG- (Fig. 2A) or vector alone (not shown). Whole cell extracts from these strains incubated with crude yeast cell lysate supernatants containing expressed gag p55 generated immunoreactive bands of 17 and 24 kDa, but only for pPRT (Fig. 2B). The 17- and 24-kDa products comigrated with authentic HIV-1 gag antigens p17 and p24, suggesting that cleavage at the correct point within the polyprotein had occurred. To determine that the expressed HIV-1 protease was responsible for the observed gag p55 cleavages, the protease containing a mutation of the highly conserved Asp-25 to Asn was expressed as above and shown not to exhibit any activity against gag p55 (Fig. 2).

Protein purification of the active enzyme was initially monitored by comigration of the antigen and gag p55 cleavage activity during column chromatography. At pH 7.8 this activity was not retained by a DEAE-Sephadex column, resulting in a substantial purification (Table I). Coincident with the appearance of p55 cleavage activity and antigen was a peptide hydrolyase activity detected with the substrate having the sequence Val-Ser-Gln-Asn-Tyr-Pro-Ile-Val, corresponding to the gag p17/p24 cleavage site between Tyr-132 and Pro-133 in gag p55 (Fig. 3A). Treatment of the peptide with active fractions from the DEAE eluate resulted in a time-dependent loss of substrate upon high performance liquid chromatography analysis, and a parallel increase in two new peaks. The new peaks were shown to comigrate with the chemically synthesized anticipated products, Val-Ser-Gln-Asn-Tyr and Pro-Ile-Val, confirming the cleavage point within the peptide. The peptide hydrolysis was not detectable in crude lysates, possibly due to binding or modification of the peptide by contaminating proteins. When the cells from the control strain containing pATG- or cells from the strain containing...
Fig. 3. DEAE-Sephadex purification of HIV [Ala-1] protease and Asn-25 mutant. A, isocratic elution profiles of absorbance at 280 nm (solid lines) and gag 128–135 peptide hydrolysis activity (dashed lines) for chromatography of supernatants from homogenates of 6 g of cells transformed with either pPRT (upper curves) or pPRT (Asn) (lower curves). Maximal peptide hydrolysis activity eluted with the pPRT sample was 41 nmol/min/ml (upper curve). No peptide hydrolysis activity could be detected with the Asn-25 mutant (lower curves). B, immunoblot analysis of protease antigen from the Asn-25 elution of DEAE-Sephadex, analyses as in Fig. 1A. Samples for each lane contained 10 μl from fractions indicated at the top of the blot. Fractions not shown here showed no antigen. Asterisk indicates the lane containing synthetic protease (0.02 μg).

pPRT (Asn-25) were lysed in an identical way and passed through the DEAE column, no peptide hydrolase activity was detectable (Fig. 3A). In trials of CM-Sephadex and mono S column purification steps, the peptide hydrolase activity was found to comigrate with gag p05 cleavage activity and protease antigen. Thus, the peptide hydrolysis was used to quantify purification steps and characterize the purified activity.

The peptide hydrolysis assay with high performance liquid chromatography detection of products typically gave signal-to-noise ratios greater than 100 for samples from the DEAE-Sephadex step. Although the Asn-25 mutant was observed on immunoblots to migrate in the same position as active protease (Fig. 3B) and the level of expression appeared to be similar (Fig. 2A) no detectable activity was observed. Thus, the residual activity is less than 1% that observed with the parent construct.

The majority of the purification of the active enzyme occurred in the first two steps. The final step (mono S) was used to remove trace high molecular mass contaminants and a protein (possibly lysozyme) of approximately 14 kDa. Recovery of activity was 7%, with an overall purification of 254-fold. Similar chromatography steps with smaller amounts of starting cell paste resulted in dramatic losses in activity, suggesting some nonspecific adsorption of the protease to the chromatography supports or glassware.

A typical pattern of co-eluting peptide hydrolysis activity and protease antigen from the mono S column is shown in Fig. 4 (compare panels A and B). Panel C shows a silver staining band present in these fractions at 10 kDa (lanes 16–24) which comigrated with the synthetic protease (lane 4). A portion of the activity eluted from the mono S column appears greater than 90% pure. Coomassie Brilliant Blue-stained gels also showed the band at 10 kDa, but no other impurities were observed. The fraction of highest purity (lane 17) from the mono S step was used to characterize the activity of the HIV-1 protease further. Gas phase sequencing of a sample at this stage demonstrated that the NH2-terminal Met (introduced to generate the initiation codon) had been removed and that Ala was the first amino acid. The subsequent 14 cycles of sequencing confirmed the anticipated sequence and was consistent with a high degree of purity.

The NH2 terminus of Met-Ala was changed to Met-Pro by site-directed mutagenesis, since Pro is the actual NH2-terminal amino acid encoded by the viral genome. Elution patterns

![Graph](image-url)
HIV Protease Expression, Purification, and Characterization

of protein and peptide hydrolysis activity were not significantly different from those observed with the [Ala-1] protease. Sequencing demonstrated the loss of the NH2-terminal Met in this case also.

At concentrations above 0.1 mg/ml the purified protein was found to be moderately stable to freeze-thaw cycles and storage at −70 °C. Stability studies with partially purified preparations showed rapid loss of activity except in the presence of 10% glycerol. The less purified DEAE run-through fractions appeared more stable at 4 °C or frozen than the more highly purified fractions.

The protease peptide hydrolysis activity was greatest at low pH with a maximum activity at pH 5.5, although activity was high from pH 4.0 to 6.5 (Table II). The activity did not appear to be sensitive to the nature of the buffers utilized. A similar pH dependence was observed for cleavage of the natural substrate gag p55 (data not shown).

Peptide hydrolysis activity obeyed simple Michaelis-Menton kinetics at pH 5.5 and a $K_m$ of 2.2 mM and $k_{cat}$ of 240 min$^{-1}$ (based on a catalytic unit of a 22-kDa dimer) were determined. The minimal peptide substrate based upon the gag p17/gag p24 cleavage site was determined from the kinetics of hydrolysis of a series of shorter peptides (Table III). At least in the case of this amino acid sequence, a minimum of 7 amino acids are required for efficient catalysis. A much longer peptide (His-Ser-Gln-Val-Ser-Gln-Asn-Tyr-Pro-Ile-Val) was tested at 1.0 and 0.5 mM and found to be hydrolyzed no more readily than the octamers shown in Table III.

The effect of the aspartyl protease inhibitor pepstatin on peptide hydrolase activity was examined with 2 and 6 mM substrate, Val-Ser-Gln-Asn-Tyr-Pro-Ile-Val. The Dixon plot shown in Fig. 5 revealed a $K_i$ of 1.1 μM. Replots of this and additional determinations showed that the inhibition was not entirely competitive, being partially noncompetitive.

Size estimates of the protease were performed by gel filtration on Sephadex G-75 after either the DEAE or mono S steps. In both cases the peak of activity eluted just after the elution position of an $\alpha$-chymotrypsinogen (25 kDa) standard, giving a Stokes radius corresponding to a 22-kDa protein. Recovery of activity when gel filtration was used after the DEAE step was 85%; the purification was minimal. When the more pure mono S pool was eluted, recovery of activity was much lower (7%) although the elution position did not change. From these data, it appears likely that the HIV-1 protease is active as a dimer.

To further examine the structure of the purified protein, circular dichroism spectra were recorded and are shown in Fig. 6. Simulation of the spectra obtained (27) produced an estimate of the secondary structures for the protease: 65% $\alpha$-helix, 75% $\beta$-sheet, 13% turn, and 5.5% “random.” These data were obtained with the protein having NH2-terminal proline, since larger quantities were available.

**DISCUSSION**

HIV-1 was postulated to encode a protease essential for viral replication based on homology with other retroviruses. The active form was anticipated to result from cleavage between residues 67 and 68, and 167 and 168 of the pol open reading frame to generate the 99-residue protein. The expression described here of that sequence, or the NH2-terminal Ala-1 sequence, resulted in an active 99-residue protease,
confirming a recent observation (14) that the flanking sequences initially present following translation of the viral polypeptide are not necessary for folding to an active form. Reports have appeared describing the expression of active enzyme from larger constructs (12, 13), including one in which the NH2 terminus of a proteolytically processed product was shown to begin as predicted, at Pro-68 of the pol open reading frame (12). It has previously not been determined that in those preparations the processed 99-residue protein was the active form.

A 10-kDa protein isolated from radiolabeled HIV-1 virions yielded NH2-terminal sequence information consistent with it being the HIV-1 protease, although no activity was detected (16). Other attempts at purification from microbial expression systems (15) have resulted in only partially pure preparations. The purification described here begins with and maintains the enzyme in a soluble state, an approach most likely to maintain native structure.

Characterization of the HIV-1 protease as an aspartic protease has been reported by mutation of the Asp-25 to Ala and loss of autoprocessing of a dimer or larger aggregate, assuming that the protease was not carried through the column by contaminating proteins.

The structural characterization of the purified HIV-1 protease has been reported by mutation of the Asp-25 to Ala and loss of autoprocessing of a dimer or larger aggregate, assuming that the protease was not carried through the column by contaminating proteins.

The structural characterization of the purified HIV-1 protease has been reported by mutation of the Asp-25 to Ala and loss of autoprocessing of a dimer or larger aggregate, assuming that the protease was not carried through the column by contaminating proteins.