PROTEINS: Structure, Function, and Genetics 6:139–154 (1989)

Characterization and Autoprocessing of Precursor and Mature Forms of Human Immunodeficiency Virus Type 1 (HIV 1) Protease Purified From *Escherichia coli*

James E. Strickler,[1] Joselina Gorniak,[1] Brian Dayton,[2] Thomas Meek,[2] Michael Moore,[3] Victoria Magaard,[3] John Malinowski,[4] and Christine Debouck[4]

Departments of [1]Macromolecular Sciences, [2]Medicinal Chemistry, [3]Peptide Chemistry, and [4]Molecular Genetics, Smith, Kline and French Laboratories, P.O. Box 1539, King of Prussia, Pennsylvania 19406

ABSTRACT **A recombinant plasmid encompassing the human immunodeficiency virus type 1 (HIV 1) protease coding sequence and flanking regions (Ala-13 to Gly-185 of the *pol* open reading frame) has been expressed in two distinct strains of *Escherichia coli*, AR58 and AR68. In the first strain, AR58, the primary translation product, a 25 kilodalton (kDa) precursor protein, is short-lived and rapidly processes itself to the 11 kDa mature protease in vivo. In the second strain, AR68, the 25 kDa species is only partially processed, and it, a 13 kDa intermediate, and the mature 11 kDa enzyme accumulate at a ratio of 3:4.5:2.5, respectively. The 11 kDa mature protease from AR58 and the 25 kDa precursor from AR68 have been purified to homogeneity. The yield of 11 kDa enzyme from AR58 is approximately 0.02 mg/g wet weight of *E. coli* cell pellet. The protease has both the expected NH_2- and COOH-terminal sequences. The yield of 25 kDa enzyme from AR68 is approximately 0.1 mg/g wet weight of *E. coli* cell pellet. In vitro, the 25 kDa precursor enzyme rapidly ($t_{1/2} \cong$ 9 min) processes itself into a species with a mass of ~13kDa and a species with a mass of ~11 kDa. Both of these latter species can be separated by RP-HPLC, have the NH_2-terminal sequence expected for the mature protease, and are active. The 11 kDa enzyme from AR58 comigrates with the 11 kDa enzyme from AR68 on RP-HPLC and SDS polyacrylamide gel electrophoresis. On extended incubation at 4°C at either neutral or acidic pH all species of the protein exhibit further autodegradation at defined sequences. The availability of the mature, 11 kDa enzyme and the 25 kDa precursor will allow biochemical and physical studies on this critical viral enzyme.**

Key words: retrovirus, bacterial expression, high-performance liquid chromatography, NH_2- and COOH-terminal sequence analysis, k_{cat}

INTRODUCTION

The human immunodeficiency virus type 1 (HIV 1) has the typical genome structure of retroviruses, i.e., 5′-*gag-pol-env*-3′.[1–4] As with other retroviruses, the primary translation products of the *gag–pol* region are polyprotein precursors requiring proteolytic processing to yield the mature protein species. This processing is carried out by a virally encoded protease, which is itself contained within one of the polyprotein products. Thus, in HIV 1, the protease is responsible for the maturation/processing of $Pr55^{gag}$ into the capsid and nucleocapsid proteins $p17^{gag}$, $p24^{gag}$, $p7^{gag}$, and $p9^{gag}$; and $Pr160^{gag\text{-}pol}$ into $p17^{gag}$, $p24^{gag}$, the protease, a reverse transcriptase, and an endonuclease. This processing has been shown to be essential for the development of infectious viral particles.[5]

The protease coding sequence begins with residue 69 of the *pol* open reading frame and extends through residue 167 yielding a 99 residue polypeptide.[6–9] This sequence is bounded by consensus protease cleavage sites having the form Ser/Thr-Xxx-Yyy-Tyr/Phe-Pro, with hydrolysis of the peptide bond occurring between the aromatic residue and the proline.[10] Cleavage at the consensus site found between residues 164 and 168 of the *pol* reading frame yields the mature NH_2-terminus of reverse transcriptase[11,12] as well as the presumptive

Received March 1, 1989; revision accepted July 7, 1989.

Address reprint requests to James E. Strickler, Department of Macromolecular Sciences, Smith, Kline and French Labs., P.O. Box 1539, King of Prussia, PA 19406-0939.

Abbreviations used: HIV 1, human immunodeficiency virus type 1; kDa, kilodalton; RP-HPLC, reverse-phase high-performance liquid chromatography; SDS, sodium dodecyl sulfate; $11kDa^{AR58}$, the 11 kDa HIV protease isolated from *E. coli* strain AR58; $25kDa^{AR68}$, the 25 kDa protease precursor isolated from *E. coli* strain AR68; DTT, dithiothreitol; PMSF, phenylmethylsulfonylfluoride; TFA, trifluoroacetic acid; HOAc, acetic acid; PTH-amino acid, phenylthiohydantoin amino acid derivative; OPA, *o*-phthaldialdehyde; ODS, octadecylsilane; MES, 2-(*N*-morpholino)-ethanesulfonic acid; HP-SEC, high-performance size-exclusion chromatography.

COOH-terminus of the protease. Induction in bacteria of plasmids containing this coding sequence, or chemical synthesis of the polypeptide deduced from it, resulted in the production of active protease.[6,7,13–20] The bacterially expressed enzyme exhibited proteolytic activity toward natural substrates in vivo and in vitro.[5–7,13,15,20,21] Both the chemically synthesized enzyme and the recombinant enzyme were capable of cleaving oligopeptide substrates containing the consensus cleavage site.[18–23] Based on homologies with other aspartyl-proteases, the HIV 1 protease has been proposed to belong to this class of enzymes.[24] The observation that pepstatin inhibits either the viral, recombinant, or chemically synthesized enzyme is consistent with this proposal.[5,14,18,20,21] Mutation of the aspartyl residue present at position 25 in the presumed active site Asp-Thr-Gly sequence to either alanine, asparagine, threonine, histidine, or tyrosine results in inactivation of the protease in vivo,[5,15,25,26] further supporting the proposal. Molecular modeling of the HIV 1 protease based on the structure of the aspartyl proteases, penicillopepsin, and endothiapepsin, suggested that the enzyme needs to be a homodimer in order to form the active site.[27] This hypothesis has been confirmed by gel filtration and ultracentrifugal studies and by solution of the three-dimensional structure of the bacterially expressed enzyme.[14,20,21,28]

The central role this enzyme plays in the life cycle of the AIDS virus makes it a logical therapeutic target. Indeed, the specificity of the enzyme for the consensus sequence Ser/Thr-Xxx-Yyy-Tyr/Phe-Pro makes it a logical target for drug development. This sequence occurs once in the *gag* open reading frame (between $p17^{gag}$ and $p24^{gag}$) and twice in the *pol* open reading frame (at the beginning and end of the protease). Maturation of $Pr55^{gag}$ and $Pr160^{gag-pol}$ requires the viral protease to process at four other sites showing little or no homology to either this consensus sequence or to each other. In general these cleavages occur between two hydrophobic residues, e.g. Leu-Ala, at the COOH-terminus of $p24^{gag}$ or Leu-Phe, at the junction of the reverse transcriptase and the endonuclease.[11] Darke et al.[19] have shown that oligopeptides containing all of the known processing sites of the authentic viral enzyme can be cleaved by chemically synthesized protease.

In order to study the processing of the protease, its detailed molecular structure and its mechanism of action, we have subcloned and expressed a precursor form of the enzyme having both NH_2- and COOH-terminal extensions in two different strains of *E. coli*. In one strain the precursor accumulates, while in the other strain it rapidly autoprocesses to the mature 11 kDa form in vivo. The availability of both forms of the enzyme will allow us to investigate structural constraints on the activity of the protease. Such constraints may have implications in the design of specific inhibitors, since the initial translation product containing the protease is a 160 kDa precursor. In this paper, we describe the purification and protein chemical characterization of both species and the in vitro autoprocessing of the precursor to yield the mature protease. Enzymological characterization of the purified, 11 kDa in vivo-processed, mature protease has been described elsewhere.[21]

MATERIAL AND METHODS

Expression

HIV protease was produced in *Escherichia coli* using the PRO4 expression vector which encodes a protease precursor extending from Ala-13 to Gly-185 of the *pol* open reading frame fused at its NH_2-terminus to a portion of a bacterial protein.[6] The *E. coli* strains used were AR58[29] [an N99 derivative with *gal*E::Tn10 (λkil^-, *c*I857 $\triangle$H1)] and a heat-shock deficient strain, AR68[30] (a CAG456[31] derivative with *gal*E::Tn10 (λcI857 $\triangle$Bam$\triangle$H1); CAG456 is a K-12 derivative with lac_{am},trp_{am}, pho_{am},$sup\mathrm{C}^{ts}$, *htp*R, str^r). Protease production was obtained by heat induction of the P_L promoter.[32] In *E. coli* strain AR58, induction of the PRO4 plasmid did not result in the accumulation of the precursor, but instead led to the appearance of a protein with an apparent molecular weight of 11,000, the expected molecular weight of mature protease.[8] This probably resulted from autocatalytic processing of the precursor in this strain of *E. coli* as we have previously described for the nalidixic acid induction of PRO4 in *E. coli* strain AR120.[6] In *E. coli* strain AR68, induction of the PRO4 plasmid resulted in the accumulation of protease species with apparent molecular weights of 25,000 (the expected molecular weight of the precursor), 13,000, and 11,000.

Purification

The recombinant HIV protease expressed in *E. coli* strain AR58, $11kDa^{AR58}$, was purified as follows. An *E. coli* cell pellet (~50 g wet weight) was suspended in a buffer consisting of 50 mM Tris-HCl, pH 7.5, 1 mM dithiothreitol (DTT), 1 mM EDTA, and 1 mM phenylmethylsulfonylfluoride (PMSF) at a ratio of ~1 g cells/2 ml buffer. In order to disrupt the cells, the suspension was sonicated at ~70% power using a 1" horn mounted in a Heat Systems Ultrasonicator Model W-385 (5–10 15 sec bursts with 30 sec cooling between each). The sonicate was centrifuged in a Sorvall RC-5B superspeed centrifuge at 20,000 *g* for 15 min in order to obtain a clarified supernatant. This supernatant solution was brought to 40% of saturation by the addition of dry ammonium sulfate at room temperature. The resulting suspension was stirred for an additional 30 min at 4°C before centrifugation at 20,000 *g* for 15 min. The supernatant solution was discarded and

the precipitate was redissolved in a minimal volume of a buffer consisting of 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 0.1 mM each of DTT and EDTA. Remaining insoluble material was removed by centrifugation as above and the supernatant fluid was loaded, usually in 5 ml aliquots, onto a 2.15 × 60 cm TSK G2000 SW column (Beckman/Altex) equilibrated in the same buffer at 4 ml/min. The absorbance of the effluent from the column was monitored at 280 nm and fractions of 1 min were collected. Aliquots of each fraction were examined by SDS-polyacrylamide gel electrophoresis and assayed for protease activity. The peak fractions were pooled, brought to 40% glycerol and stored at −20°C.

All of the protease species expressed in *E. coli* strain AR68 were found to be insoluble upon lysis of the cells. Cell pellets were suspended and lysed as described above. The sonicate was centrifuged at 20,000 *g* for 15 min at 4°C and the supernatant solution was discarded. The precipitate was then suspended in 1 M urea, 20 mM Tris-HCl, pH 7.5,1 mM DTT, and 1 mM EDTA. This suspension was centrifuged as described and the supernatant solution discarded. Finally, the precursor and the other protease species were extracted from the precipitate with 8 M urea, 20 mM DTT, 20 mM Tris-HCl, pH 7.5. The extract was acidified by the addition of 10% trifluoroacetic acid (TFA) to a final concentration of 0.1% and the resulting precipitate was removed by centrifugation as above. The supernatant solution was isolated and chromatographed on a Vydac C_{18} column (2.2 × 25 cm, Bodman Chemical Co.) The column was equilibrated in 24% CH_3CN/0.05% TFA at a flow rate of 10 ml/min. Elution of bound proteins was accomplished with a linear gradient to 48% CH_3CN/0.05% TFA over 60 min. The absorbance of the column effluent was monitored at 280 nm and 1 min fractions were collected. Aliquots of each fraction were dried in a Speed-Vac concentrator in the presence of 0.005% Triton X-100 and dissolved in 0.2% HOAc. The samples were analyzed both for protease activity and by SDS-polyacrylamide gel electrophoresis. Protease-containing peak fractions were pooled and lyophilized. The 25 kDa pool was redissolved in 0.2% HOAc and subjected to a second round of RP-HPLC in order to eliminate contaminating 11 and 13 kDa protease species (see Results).

Analytical Methods

Protein determinations were made by a modification[33] of the Coomassie dye binding assay of Bradford[34] using BSA as a standard. SDS-polyacrylamide gel electrophoresis analysis was performed after the method of Laemmli[35] on 0.4 mm thick 10 cm long minigels (Aquebogue, NY). The gels were stained first with Coomassie brilliant blue R250 and then by the silver-staining method of Merril et al.[36] For immunoblot analysis, proteins were electrophoretically transferred from unstained SDS-polyacrylamide gels onto polyvinylidene difluoride (PVDF) membranes (Immobilon, Millipore Corp.) The resultant blots were blocked with 0.6% gelatin and the protease localized using a protease-specific rabbit polyclonal antiserum[6] and a biotin-avidin based staining system (Vectastain, Vector Labs, Inc.). Immunodotblots were prepared by drawing a grid on a piece of PVDF filter and applying 2 μl aliquots of fractions within the grid. The dotblots were air dried and then processed as described above.

Analytical RP-HPLC was performed in either of two systems. The first system, used for identification and quantification, consisted of a Vydac C_{18} column (4.6 × 250 mm) mounted in a Beckman System Gold LC equipped with an IBM PC/AT for data collection and analysis. The absorbance at 214 nm of the effluent was monitored with a Hewlett-Packard HP1050 Series UV/Vis detector. The column was equilibrated at 1 ml/min in 24% CH_3CN/0.05% TFA and developed with a linear gradient to 48% CH_3CN/0.05% TFA over 20 min. A protease calibration curve ($R = 0.99$) was generated by injecting aliquots of a protease solution for which the concentration had been determined by amino acid analysis. The second system, used for microanalytical and preparative work, was a microbore HPLC system consisting of a Brownlee model G micropump, a homemade 50 μl dynamic mixer, a Rheodyne model 7125 injection valve, a Brownlee RP-300 column (1 × 250 mm), and a Hewlett-Packard model 1040A Diode Array Detector. The system was operated at a flow rate of 50 μl/min. Gradient details are given in the figure legends.

Amino Acid and Sequence Analysis

Amino acid analysis was performed by gas-phase hydrolysis of samples at 110°C for 16–18 hr using constant boiling HCl (Pierce Chemical Co.). Hydrolysates were analyzed on a Beckman 6300 amino acid analyzer equipped with an SICA model 7000S integrator (Scientific Instruments Corporation of America) which was used for peak identification and quantification.

Polypeptides were subjected to automated Edman degradation in an Applied Biosystems Model 470A Gas-Phase Sequenator. The resulting PTH-amino acids were analyzed by RP-HPLC using a Beckman System Gold LC equipped with an IBM PC/AT for data collection and analysis. The PTH-amino acids were separated on a Beckman Ultrasphere ODS column (4.6 × 250 mm) operating at 1 ml/min and heated to 40°C. The column was equilibrated in 24% CH_3CN/20 mM sodium acetate (NaAc), pH 4.1 (A buffer) and developed with a linear gradient to 50% B in A (B buffer was 60% CH_3CN/50 mM NaAc, pH 5.5) over 15 min. The column was held at 50% B for 3 min before stepping back to 100% A. The effluent

from the column was passed through a Kratos Model 783 UV/Vis detector programmed to monitor at 269 nm from 0 to 11.1 min in the run, switch to 323 nm from 11.1 to 13.7 min and then switch back to 269 nm for the duration of the run. PTH-amino acid derivatives were identified and quantitated by comparison to a standard mixture of PTH-amino acids (Pierce Chemical Co.).

Carboxy-terminal sequencing was performed by digestion of samples with either carboxypeptidase Y or carboxypeptidase P for varying lengths of time at room temperature. Released amino acids were quantitated by derivatization with *o*-phthaldialdehyde (Pierce Opa reagent). The derivatized amino acids were analyzed by a modification of the method of Jones and Gilligan.[37] Briefly, the OPA-amino acids were separated on a Beckman Ultrasphere 3μ-ODS column (4.6 × 75 mm) equipped with a Brownlee Newguard RP-18 guard column (3.2 × 15 mm). The effluent from the column was monitored with a McPherson model 750 spectrofluorometer equipped with the high sensitivity attachment. Excitation was at 230 nm and emission was monitored through a 500 nm cutoff filter. The output from the fluorometer was digitized using a Beckman Model 406 Analog Interface Module and collected on an IBM PC/AT. Data reduction and analysis were performed using Beckman System Gold software.

Protease Assay

The assay was essentially as described.[21] Briefly, aliquots of fractions were incubated with the nonapeptide substrate, Ac-Arg-Ala-Ser-Gln-Asn-Tyr-Pro-Val-Val-NH_2, for varying lengths of time at 37°C in a buffer consisting of 50 mM MES, pH 6.0, 200 mM NaCl, 1 mM DTT and EDTA, and 0.1% Triton X-100. The total reaction volume was 9 μl, and the assay was stopped by the addition of 1 μl 10% TFA. The samples were then diluted to 25 μl with 0.1% TFA for injection of 5 μl onto the analytical system. Separation of the product from the substrate was accomplished by RP-HPLC on a Brownlee RP300 cartridge (4.6 × 100 mm) mounted in a Beckman System Gold LC. The column was equilibrated at 1 ml/min in 0.05% TFA and developed with a linear gradient to 18% CH_3CN/0.05% TFA over 15 min. The column effluent was monitored at 214 nm and the data were collected and analyzed on an IBM PC/AT using Beckman System Gold software. The putative product peak from one such assay was collected and subjected to amino acid analysis and COOH-terminal sequencing in order to confirm its identity. The percent conversion (product peak area)/[(product peak area) + (residual substrate peak area)]) was used to calculate units of activity (1 mU = 1 nmol of product produced/min).

RESULTS

Purification of Soluble, Mature Protease From *E. coli* Strain AR58

Heat induction of the PRO4 plasmid in *E. coli* strain AR58 did not result in the accumulation of the expected 25 kDa precursor. Instead a protein with an apparent molecular weight of 11,000 was produced at a level of ~0.04% of the total cellular protein. This result exactly mirrored our experiences with the nalidixic acid induction of this plasmid in *E. coli* strain AR120.[6] The fact that this vector/host combination produced an enzyme that was soluble and exhibited activity towards a polypeptide substrate derived from $Pr55^{gag}$,[21] made it an excellent source for purification of the protease. The soluble fraction of lysed, induced cells was isolated and brought to 40% of saturation with dry ammonium sulfate in order to concentrate the protease and effect a partial purification. By immunoblot analysis, ≥80% of the protease precipitated under these conditions (Fig. 1A). The precipitate was redissolved as described in Materials and Methods and the clarified sample was subjected to preparative high-performance size-exclusion chromatography (HP-SEC) on a TSK G2000 SW column (Fig. 1B). All of the activity and immunoreactive protein eluted in a single peak between 44 and 47 min. Elution in this position is consistent with an apparent molecular weight of ~5100, much lower than either the protease monomer molecular weight of 11,000 or the expected dimer molecular weight of 22,000. This elution behavior presumably results from an interaction of the protease with the silica-based support, since the apparent molecular weight of the protein eluting in this position, as determined by SDS polyacrylamide gel electrophoresis, is ~11,000 (Fig. 1C). Furthermore, in separate studies, the $11kDa^{AR58}$ protease behaved as a dimer both by Sephacryl S-200 chromatography and by glycerol density ultracentrifugation.[21] The retardation obtained with the TSK-column probably aids in the purification of the enzyme due to the elution of contaminating bacterial proteins in a more normal fashion. SDS-polyacrylamide gel electrophoresis analysis of the fractions containing peak protease activity indicates that the protease is ≥90% pure at this stage. In order to desalt the protein, the pooled protease peak was subjected to reverse-phase HPLC (RP-HPLC) (Fig. 2). Protein recovery from this support ranged from 80 to 100% with full recovery of enzymatic activity (data not shown). RP-HPLC was not used as a routine purification step for the $11kDa^{AR58}$ protease in order to avoid potential denaturation; however, protease prepared by this method was quantitated by amino acid analysis and the quantified sample used to calibrate an analytical RP-HPLC system. RP-HPLC was also used to prepare samples for structural studies.

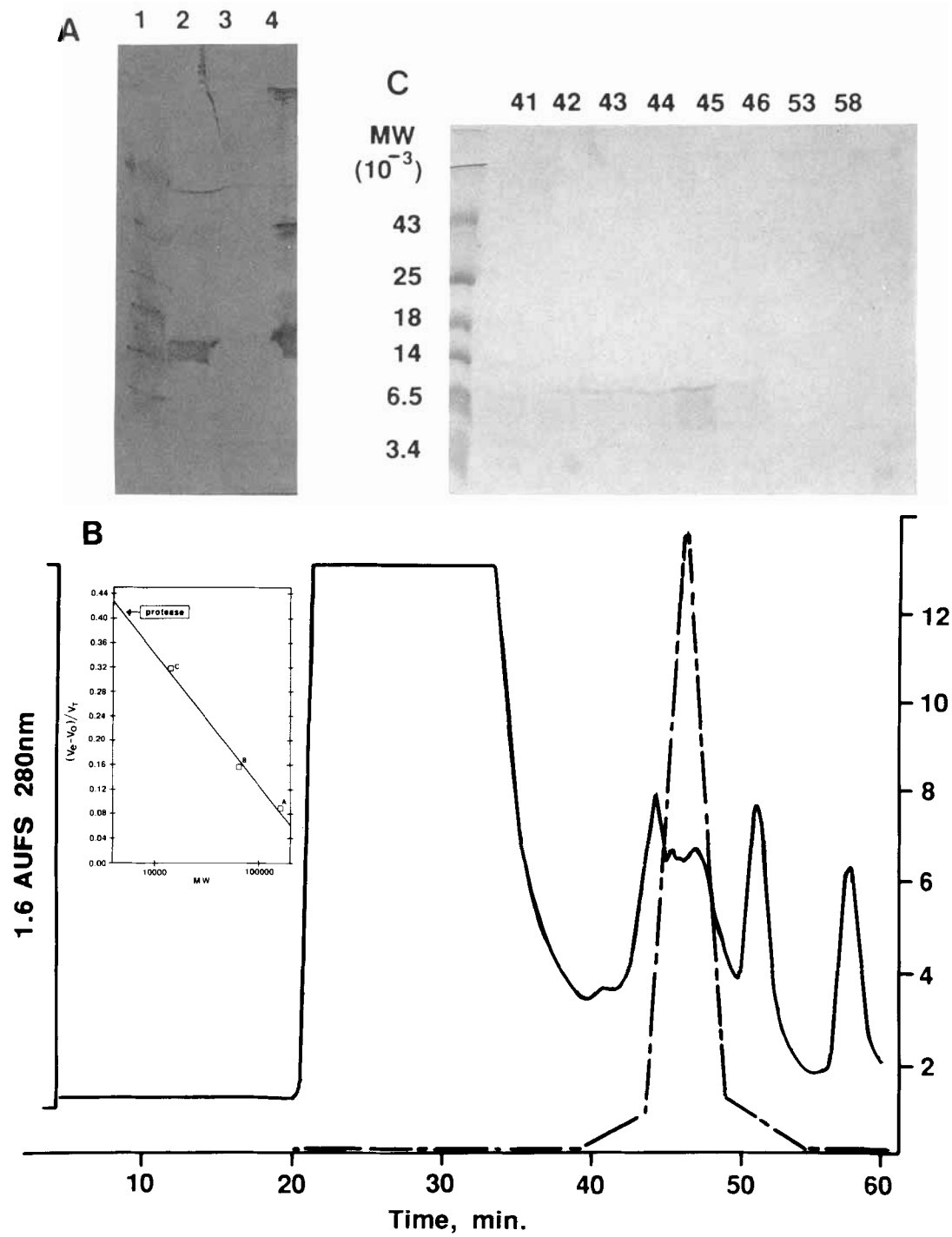

Fig. 1. Purification of mature 11 kDa HIV protease expressed in *E. coli* strain AR58. **(A)** SDS-polyacrylamide gel electrophoresis immunoblot analysis of proteins derived from the $(NH_4)_2SO_4$ fractionation of induced *E. coli* strain AR58 harboring the PRO4 plasmid as described in Materials and Methods. Lane 1 is prestained markers, lane 2 is an aliquot of the crude lysate soluble fraction, lane 3 is an aliquot of the 40% $(NH_4)_2SO_4$ precipitate, and lane 4 is an aliquot of the 40% $(NH_4)_2SO_4$ soluble fraction. **(B)** TSK G2000SW HP-SEC of the 40% $(NH_4)_2SO_4$ precipitate. The precipitate derived from 50 g of starting material was redissolved in 25 ml of 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 0.1 mM DTT, 0.1 mM EDTA. Aliquots (5 ml) were chromatographed on a 2.15 × 60 cm TSK G2000SW column equilibrated in the same buffer at 4 ml/min. The solid line shows the elution of 280 nm absorbing material from the column. The dashed line shows the elution of specific HIV protease-related peptidolytic activity. The inset shows the elution position of the protease peak with respect to marker proteins of known molecular size (A = aldolase, B = bovine serum albumin, and C = ribonuclease A). **(C)** SDS-polyacrylamide gel electrophoresis analysis of fractions corresponding to the protease activity peak. The gel was stained with Coomassie brilliant blue R250. Numbers above the wells indicate the individual fraction numbers analyzed. Molecular weight markers are shown in lane 1. They are ovalbumin, 43,000; chymotrypsinogen, 25,000; β-lactoglubulin, 18,400, lysozyme, 14,700; aprotinin, 6,500; and insulin B-chain, 3,400.

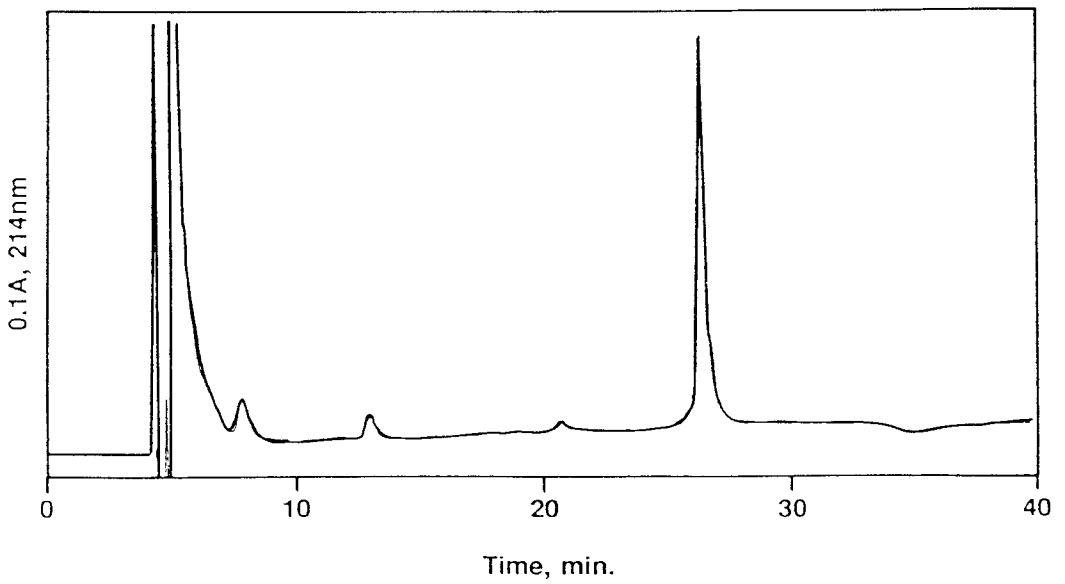

Fig. 2. Reverse-phase HPLC of 11kDa^{AR58} protease. An aliquot of the pooled TSK G2000SW protease peak was chromatographed on a Brownlee RP300, C_8 column (4.6 × 100 mm) as described in Materials and Methods under analytical methods. The column was developed with a gradient from 30 to 60% CH_3CN over 30 min. The peak fraction was collected and used for amino acid analysis, NH_2-terminal sequence analysis, or COOH-terminal sequencing.

The quantitative recovery of protease activity through this purification scheme is shown in Table I. It should be pointed out that the activity measured in the crude lysate fairly represents the actual amount of protease present. If a similar lysate is prepared from induced *E. coli* harboring the PRO4 vector containing a mutation in the protease gene, no specific peptidolytic activity can be measured.[21] The apparent low recovery at the $(NH_4)_2SO_4$ step is likely an artefact of partial inhibition of the assay. By immunoblot assay (Fig. 1A), >80% of the immunoreactive material is found in the ammonium sulfate precipitate. By immunodotblot analysis, the elution of immunoreactive protein from the HP-SEC column precisely correlates with the elution of protease activity. Our overall yield from several different isolations ranges from 0.5 to 1 mg of 11kDa^{AR58} protease/50 g *E. coli* cell pellet with a recovery of 50–70%. The fold purification is >5000. The specific activity of the protease, based on the peptidolysis assay, ranges from 3.5 to 4 × 10^4 mU/mg (35–40 μmol product min^{-1} mg^{-1}.

Chemical Characterization of Purified 11kDa^{AR58} Protease

In order to determine if the precursor is processed correctly in vivo, we performed NH_2- and COOH-terminal sequence analysis as well as amino acid composition analysis on the purified 11kDa^{AR58} enzyme. Table II shows the results of NH_2-terminal sequence analysis of the RP-HPLC purified protease and Table III shows the composition of the protein. As can be seen, the NH_2-terminus shows excellent correspondence with the sequence expected if processing occurred between the Phe and Pro within the Ser-Phe-Asn-Phe-Pro consensus sequence at residues 65–69 of the *pol* reading frame (numbering as in 1). The next consensus sequence for processing by retroviral proteases, Thr-Leu-Asn-Phe-Pro, starts 95 amino acids downstream from this site (residues 164–168 of the *pol* reading frame) and is known to give rise to the NH_2-terminus of the reverse transcriptase.[11,12] Except for Ile, the composition of the 11kDa^{AR58} protease shows good correlation with the polypeptide predicted between these consensus sites. Since the sample was only hydrolyzed for 16 to 18 hr, we expect that Ile and Val would yield lower than expected values. In order to determine if this second site was used for processing in our bacterial expression system, we performed COOH-terminal sequence analysis on the RP-HPLC-purified 11kDa^{AR58} enzyme. The results of this analysis are shown in Figure 3. The released amino acids and the kinetics of their release are consistent with the sequence -Thr-Leu-Asn-Phe-COOH. At longer digestion times, Gly, Ile, and Gln were also found, although the order of their release was less clear. These results are consistent with the sequence -Gln-Ile-Gly-(Cys)-Thr-Leu-Asn-Phe-COOH expected for cleavage at this site. In our hands, native cysteine is not recovered nor quantitated well with the OPA amino acid analysis system. We did not attempt to reduce and alkylate the protein in order to positively identify the cysteine. From these results, it appears that the 11kDa^{AR58} enzyme is equivalent to authentic HIV protease.[8,9]

TABLE I. Purification of 11-kDa Protease From *Escherichia coli* Strain AR58

Sample	Volume (ml)	Total activity (mU × 10^{-3})	Total protein (mg)	Specific activity (mU/mg)	Purification (fold)	Recovery (%)
Crude lysate	146	55.6	8300*	6.7	1	(100)
Supernatant	112	57.8	4900*	11.8	1.8	104
40% AS ppt.	25	27.8	1125*	24.7	3.7	50
HP-SEC pool†	35	38.9	1‡	38,900	5800	70

*Estimated by Coomassie dye binding assay.
†Two peak fractions (by activity) from each of 5 runs pooled (5 ml of 40% ammonium sulfate precipitate/HP-SEC run).
‡Determined by quantitative RP-HPLC assay.

TABLE II. Amino Terminal Sequence Analysis of RP-HPLC Purified 11kDa^{AR58} Protease

Cycle number	Amino acid	Expected*	Residue number†
1	Pro	Pro	69
2	Gln	Gln	70
3	Ile	Ile	71
4	Thr	Thr	72
5	Leu	Leu	73
6	Trp	Trp	74
7	Gln	Gln	75
8	Arg	Arg	76
9	Pro	Pro	77
10	Leu	Leu	78
11	Val	Val	79
12	Thr	Thr	80
13	Ile	Ile	81
14	Lys	Lys	82
15	Ile	Ile	83
16	Gly	Gly	84
17	‡	Gly	85
18	‡	Gln	86
19	Leu	Leu	87
20	Lys	Lys	88
21	Glu	Glu	89
22	Ala	Ala	90
23	Leu	Leu	91
24	(Leu)	Leu	92
25	Asp	Asp	93
26	Thr	Thr	94
27	(Gly)	Gly	95
28	Ala	Ala	96
29	(Asp)	Asp	97
30	Asp	Asp	98

*Based on cleavage at the first consensus cleavage site, Ser-Phe-Asn-Phe*Pro, in the *pol* reading frame of the BH10 HTLVIIIB isolate.[38]
†Based on the *pol* open reading frame.
‡Lost due to sequencer error (cf. Table IV).

TABLE III. Amino Acid Composition of RP-HPLC Purified 11kDa^{AR58}

Amino acid	Residues/ mole	Expected*
Asp	7.0†	7
Thr	7.7	8
Ser	1.4	1
Glu	9.7	10
Pro	6.2	6
Gly	12.9	13
Ala	3.2	3
Val	5.1	6
Met	2.0	2
Ile	9.0	13
Leu	11.5	13
Tyr	0.9	1
Phe	1.9	2
His	1.1	1
Lys	5.5	6
Arg	3.5	4
Totals	88.6	95

*Based on cleavage at consensus sites as described. Does not include 2 Cys and 2 Trp.
†Data are normalized to Asp.

The purified 11kDa^{AR58} protease displayed marked instability upon storage at pH 7.5 and 4°C. We have identified two mechanisms by which activity loss can be explained. First, in a few samples of the protease in which the loss of activity occurred in 24 to 48 hr, we observed a slight decrease in the RP-HPLC retention time of the enzyme (Fig. 4). The original retention time and activity could be recovered by incubation of the enzyme in 10 mM DTT at 37°C for 10 min. This result implied that one or both cysteines in the molecule underwent a time-dependent, reversible oxidation. A sample that had reversibly lost most of its activity was subjected to SDS-polyacrylamide gel electrophoresis analysis in the presence or absence of 2-mercaptoethanol. With or without reduction, the protein migrated at a position consistent with a monomeric species (i.e., 11 kDa, data not shown). Thus, it would appear that the oxidation does not involve an intermolecular crosslink between adjacent cysteines. Second, and more frequently, the loss in activity occurred over several weeks of storage at 4°C. Using the RP-HPLC assay we found a time-dependent decrease in the peak corresponding to intact 11kDa^{AR58} protease and concomitant increases in new, much earlier eluting peptide peaks (Fig. 5). In order to determine if these peptides resulted from proteolysis of the enzyme, we isolated them and determined their NH_2-terminal sequences. The results of these analyses are shown in Table IV. The fragments are consistent with cleavages between Leu-5 and Trp-6 and between Leu-63 and Ile-64. Over an extended period (3 months, data not shown) no further breakdown

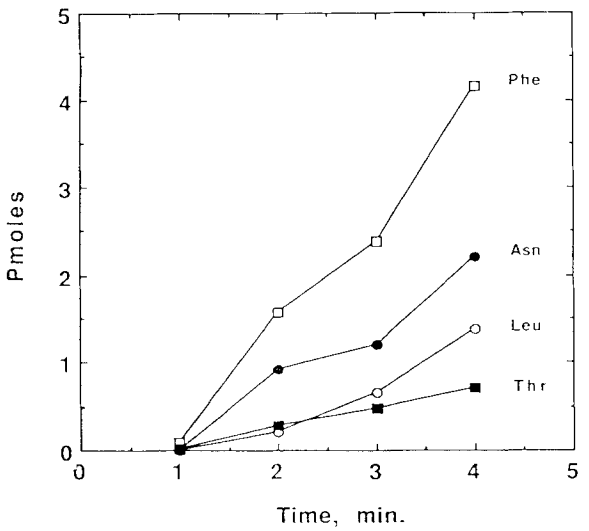

Fig. 3. Carboxy-terminal sequence analysis of RP-HPLC purified 11kDa^{AR58} protease. A lyophilized aliquot of RP-HPLC purified protease (~1 nmol) was dissolved in 10 M urea. The sample was diluted 1:1 with 100 mM NaAc, pH 6.0, and digestion started by the addition of carboxypeptidase Y at a ratio of 1:50 wt/wt (enzyme:substrate). At 0, 2, 4, 8, and 16 min, 2 μl aliquots were mixed with 8 μl of EtOH and stored on dry ice. Individual time points were removed, centrifuged, and 5 μl of the supernatant solution isolated and reacted with 5 μl of OPA reagent for 1 min at room temperature. The reaction was stopped by the addition of 10 μl 5% HOAc and 10 μl was analyzed by RP-HPLC as described in Materials and Methods. The graph shows the amounts of Phe (open squares), Asn (closed circles), Leu (open circles), and Thr (closed squares) found for each time point. The amounts of each amino acid found at 0 min of digestion have been subtracted for each time point.

was observed. The lack of additional degradation, at least partially, results from oxidation of one or both cysteines. A peak that migrates in a position consistent with the oxidized protease can be seen in Figure 5B and C (arrow). On the other hand, complete loss of the authentic peak (asterisk) was not seen suggesting that the residual, unoxidized protease may have been denatured by another mechanism leading to inactivation and resulting in no further cleavage.

Purification and Characterization of the Insoluble Protease Precursor Expressed in *E. coli* Strain AR68

Unlike heat induction of the PRO4 plasmid in *E. coli* strain AR58, induction of this plasmid in *E. coli* strain AR68 resulted in the accumulation of the 25 kDa precursor protein, a 13 kDa intermediate, and the mature 11 kDa protease. The 25 kDa species represented ~30% of the total protease-related protein while the 13 kDa species accounted for ~45% and the 11 kDa species made up the remaining 25%. The reason for the difference in the results of inducing this plasmid in these two strains is not clear. Under optimal conditions, the 25kDa^{AR68} species reached 0.2% of the total cell protein. This precursor is of interest because it can serve as an in vitro model system for the processing of viral polyproteins. On lysis of induced cells, all of the protease species were found to be insoluble; hence, they were extracted from the lysate-insoluble fraction with 8 M urea and 20 mM DTT. After acidification of the extract with TFA, the sample was chromatographed on a semipreparative Vydac C_{18} RP-HPLC column (Fig. 6A). All three species of the protease could be partially separated and pooled as highly enriched fractions. The pooled protease samples, if kept in the RP-HPLC solvents, appeared indefinitely stable at 4°C. The presence of acetonitrile appeared to completely inhibit both the proteolytic (see below) and peptidolytic activity of the enzyme. In order to eliminate cross-contamination of the different protease species, the pooled samples were made 0.005% in Triton X-100, lyophilized, redissolved in 0.2% HOAc, and rechromatographed on a second semipreparative C_{18} RP-HPLC column (Fig. 6B). The peak eluting at 25–30 min contained the 25 kDa precursor (peak 1b). This peak was collected as a pool and lyophilized in the presence of 0.005% Triton X-100. The large peak eluting around 40 min is due to Triton X-100 added to the pool prior to the lyophilization step. A typical SDS-polyacrylamide gel electrophoresis analysis of the purified 25kDa^{AR68} precursor pool is shown in the inset of Figure 6B.

Autoprocessing of the Purified 25kDa^{AR68} Precursor

Incubation of the 25kDa^{AR68} precursor at 37°C and pH 7.0 for 30 min resulted in an increase in the total activity present in the sample of ~2-fold (Fig. 7). Over the same time course there was a loss of the 25 kDa species and the appearance of polypeptides at ~13 kDa and 11 kDa. The 11 kDa species comigrated with the 11kDa^{AR58} protease on RP-HPLC (data not shown). We isolated both the 11 kDa and 13 kDa polypeptides by RP-HPLC and sequenced their NH_2-termini. Both had the sequence expected for mature HIV protease, i.e., NH_2-Pro-Gln-Ile-Thr-Leu-Trp-etc. Thus, it appears that the 25kDa^{AR68} precursor is capable of correctly processing itself to the mature form of the protease in vitro. Based on this time course, under the conditions presented, the 25 kDa precursor had a half-life of ~9 min. The isolated 13 kDa intermediate exhibited similar autoprocessing to the 11 kDa mature form in vitro (data not shown). When the in vitro-processed 11 kDa protease was stored at 4°C in 0.2% HOAc, 0.005% Triton X-100 for extended periods, further degradation was observed. Cleavage at the two sites already defined for the 11kDa^{AR58} protease were found along with one additional site. This site represents cleavage between Leu-33 and Glu-34. Presumably Glu-34 is uncharged under these conditions. The activity of this smaller species has not been examined.

Interpretation of peptidolysis assay data generated with the 25 kDa precursor could be complicated

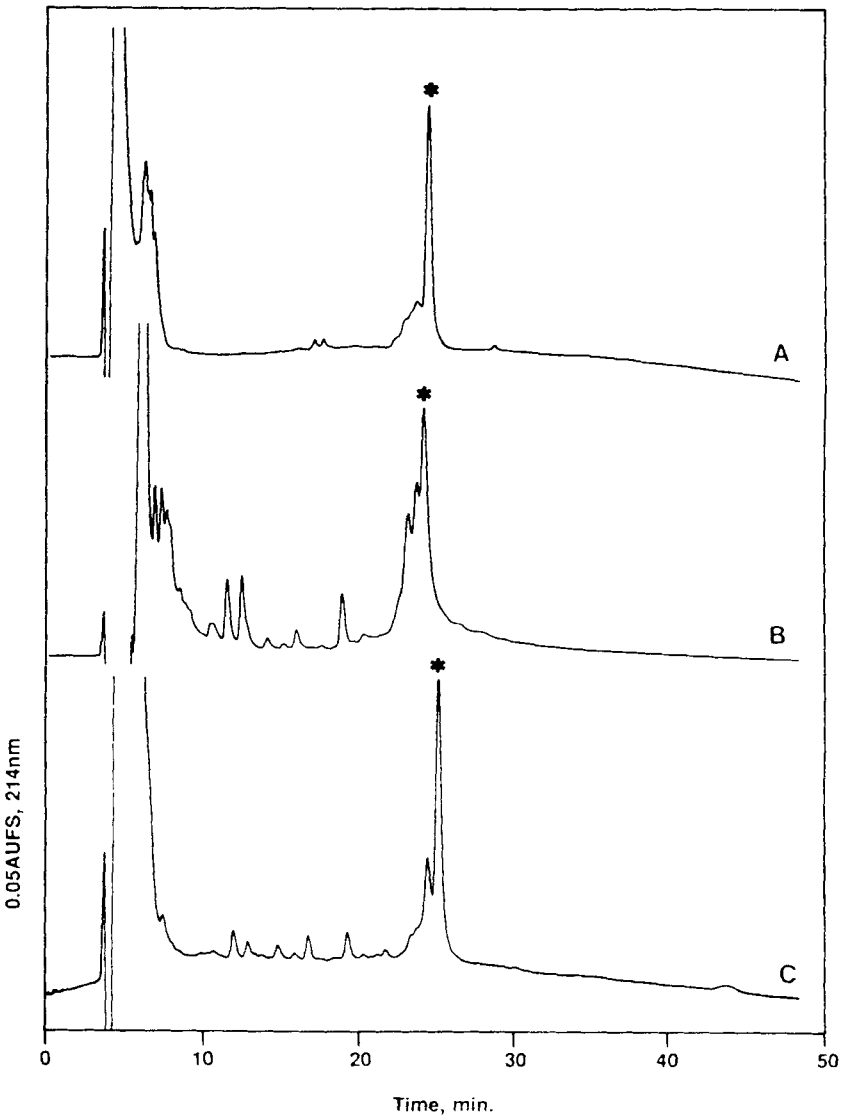

Fig. 4. Oxidation of 11kDa^{AR58} protease on storage at neutral pH at 4°C. A sample of RP-HPLC purified protease was dialyzed against 50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 0.1 mM DTT, 0.1 mM EDTA at 4°C. The dialyzed enzyme was stored at 4°C in the same buffer. The peptidolytic activity was determined immediately following dialysis and after 24 hr at 4°C. After storage, the enzyme had lost >50% of its activity. Aliquots of the protease preparation were examined by RP-HPLC. **(A)** The HPLC trace obtained with the freshly dialyzed enzyme, **(B)** is the same sample after 24 hr, and **(C)** is the sample shown in **(B)** following reduction with 10 mM DTT for 10 min at 37°C. The samples were analyzed on a Brownlee RP300 column (1 × 250 mm) equilibrated at 30% CH_3CN/0.1% TFA at 50 μl/min. The proteins were eluted with a linear gradient to 60% CH_3CN/0.1% TFA over 45 min. The asterisk indicates the normal elution position of the intact protease.

by the rapid autoprocessing of the enzyme during the course of the assay yielding several species exhibiting different kinetic parameters. In order to investigate this, we examined the protease species remaining after a typical 5 min assay on the 25kDa^{AR68} precursor by RP-HPLC and SDS polyacrylamide gel electrophoresis and found that there was little or no autoprocessing (data not shown). Apparently the >1000-fold molar excess of the peptide substrate effectively competes for the active site of the enzyme precursor. Based on the peptidolysis assay, the 11 kDa mature form and the 25 kDa precursor were found to have essentially identical K_m's for the nonapeptide substrate, 7.4 and 9 mM, respectively (this difference is probably not significant). The k_{cat}s on the other hand were quite different, ~50 s^{-1} and ~4.5 s^{-1}, respectively. In order to determine if the lower k_{cat} obtained with the 25 kDa precursor resulted from denaturation of the enzyme during RP-HPLC, we subjected a sample 11kDa^{AR58} protease to RP-HPLC, collected the protein, lyophilized it, and dissolved it in 0.2% HAc as we do with the 25kDa^{AR68} precursor. The protease thus prepared had kinetics indistinguishable from those of

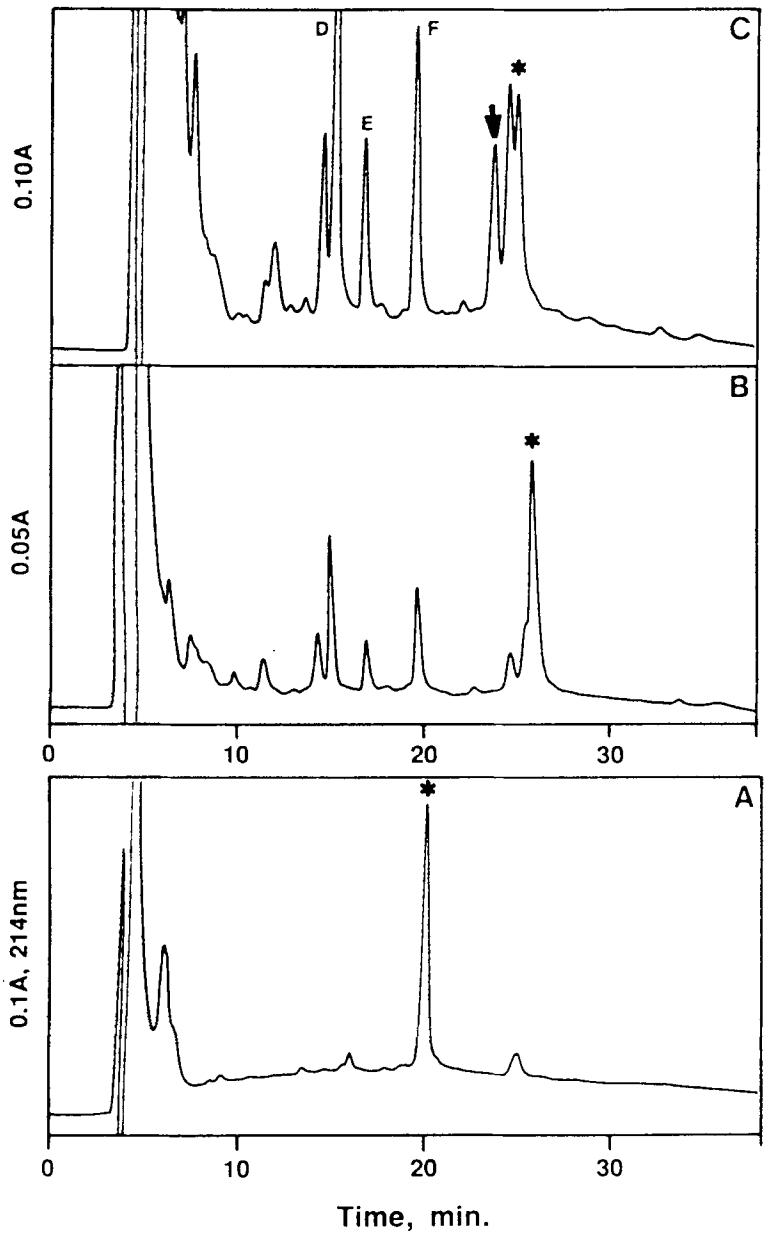

Fig. 5. The time-dependent degradation of 11kDa^{AR58} protease on storage at 4°C. An aliquot of TSK G2000SW purified protease was held at 4°C in the column buffer (see Materials and Methods) for approximately 4 months. Periodically, aliquots were removed and analyzed by RP-HPLC on a Brownlee RP300 column (1 × 250 mm). **(A)** The trace obtained with freshly prepared enzyme. The gradient was 30 to 60% CH_3CN over 30 min. **(B)** The trace obtained after storage of the enzyme for 9 days at 4°C. The gradient was 30 to 60% CH_3CN over 45 min. **(C)** The trace obtained after storage of the protease for 17 days (the gradient was the same as for B). Note: The samples shown in **A** and **B** are ~200 ng of enzyme while 400 ng was injected for **C**. The asterisk indicates the elution position of intact protease. The arrow indicates the elution position of "oxidized" protease. Peaks D, E, and F were collected for NH_2-terminal sequence analysis.

the enzyme prior to RP-HPLC; thus, it appears that at least one round of RP-HPLC and lyophilization is not sufficient to inactivate the enzyme. These data indicate that the 25 kDa precursor has intrinsically lower activity than the mature species.

DISCUSSION

We have expressed two forms of a recombinant HIV-1 protease in *Escherichia coli,* purified the proteins and chemically characterized them. The 11kDa^{AR58} protease is essentially our reference standard. It is expressed in soluble, active form and no denaturants are used in its purification (RP-HPLC is used only to prepare enzyme for structural studies). This protein has been extensively studied with regard to its substrate requirements[23] and its enzymology.[21] Because this enzyme results from in vivo processing in a foreign system, it is important

TABLE IV. NH_2-Terminal Sequence Analysis of 11kDa^{AR58} Breakdown Products

	Peak D		Peak E		Peak F	
Cycle number	Amino acid	Residue number*	Amino acid	Residue number*	Amino acid	Residue number*
1	Trp	6	Pro	1	Ile	64
2	Gln	7	Gln	2	Glu	65
3	Arg	8	Ile	3	Ile	66
4	Pro	9	Thr	4	Xxx†	67
5	Leu	10	Leu	5	Gly	68
6	Val	11	Trp	6	His	69
7	Thr	12	Gln	7	Lys	70
8	Ile	13	Arg	8	Ala	71
9	Lys	14	Pro	9	Ile	72
10	Ile	15	Leu	10	Gly	73
11	Gly	16	Val	11	Thr	74
12	Gly	17	§		Val	75
13	Gln	18			Leu	76
14	Leu	19			Val	77
15	Lys	20			Gly	78
16	Glu	21			Pro	79
17	Ala	22			Thr	80
18	Leu	23			Pro	81
19	Leu	24			Val	82
20	Asp	25			Asn	83
21	§				Ile	84
22					Ile	85
23					Gly	86
						§

*Protease sequence numbering, i.e., Pro-69 of the *pol* orf = Pro-1 of the protease.
†Expected residue is Cys.
§Not carried any further.

to prove that it displays the characteristics of the viral protease. We have shown in this paper that the 11kDa^{AR58} protease has the correct NH_2- and COOH-termini.[8,9] In our previous work, we showed that it could be inhibited by pepstatin and that it exists in solution as a dimer.[21] Thus, the 11kDa^{AR58} protease has all of the properties shown or hypothesized for the viral enzyme. We have not shown, however, that the 11kDa^{AR58} species is capable of carrying out all of the cleavages necessary for the complete maturation of Pr55^{gag} and Pr160$^{gag-pol}$. It is interesting to note in this regard that the degradation we have observed in the purified, mature protease occurs between a leucine residue and a subsequent hydrophobic residue (Trp or Ile) and is therefore similar to cleavage at the nonconsensus processing site, e.g., cleavage at the COOH-terminus of p24 between Leu and Ala or cleavage at the reverse transcriptase/endonuclease junction between Leu and Phe. Although we can not rule out the possibility that a contaminating bacterial protease present in our purified enzyme preparation was responsible for this degradation, the fact that we observed similar degradation in the 11 kDa protease isolated from *E. coli* strain AR68, which was purified in a very different fashion and therefore should not have the same contaminants as the 11kDa^{AR58} protease, makes this unlikely. Thus, these apparent autodegradations suggest that the 11kDa^{AR58} protease is capable of carrying out all of the processing steps required.

Darke et al.[20] have recently described the purification and characterization of an 11 kDa form of the enzyme expressed in *E. coli*. They inserted a methionine codon N-terminal to Pro-1 of the mature protease and a termination codon after Phe-99 in order to construct their expression vector. As with the 11kDa^{AR58} protease, the enzyme they isolated is soluble after lysis of the cells. The amount they purified was comparable to the amount we obtained from the AR58 expression system, ~1 mg/50 g cell pellet, although their overall yield was lower (~7% versus our ≥50%). They observed a loss of activity on storage as did we; however, they attributed no cause to this phenomenon. Interestingly, they observed the biggest loss in their purification scheme between their first and second steps (DEAE-Sephadex and phosphocellulose chromatography). At this stage, they carried out a concentration and dialysis step at pH 6.5. Under such conditions we would expect to see a high amount of autodegradation by the protease. It is possible that this could account for the loss they observed. Comparison of the k_{cat} of their enzyme (~4 s^{-1}) with that we obtained with the 11kDa^{AR58} enzyme (~50 s^{-1}) suggests that there may be other important differences between the two

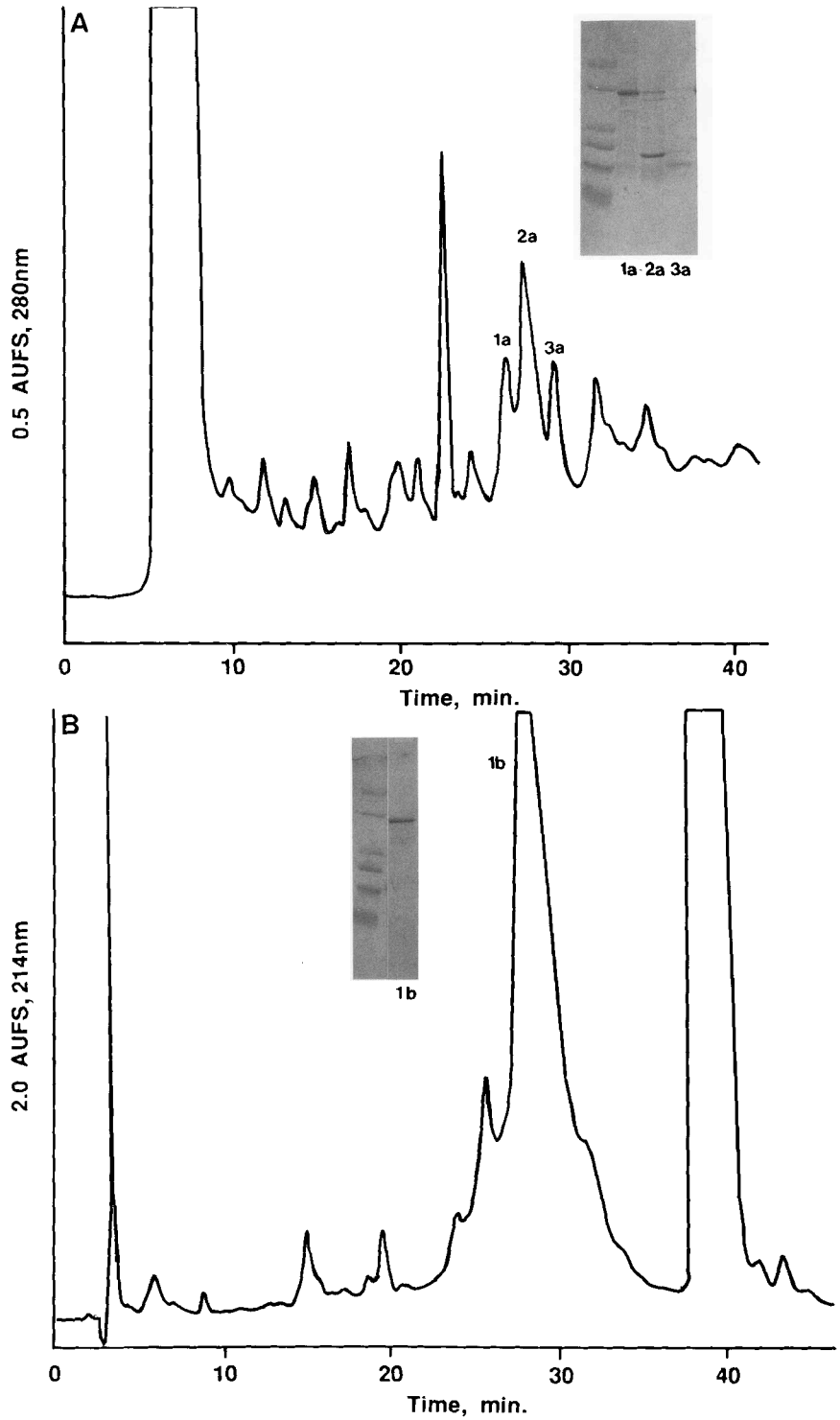

Fig. 6. Purification of the 25 kDa protease precursor from *E. coli* strain AR68. **(A)** The three protease species, extracted from the *E. coli* lysate insoluble fraction with 8 M urea and 20 mM DTT, were chromatographed on a Vydac C_{18} column (2.2 × 25 cm) as described in Materials and Methods. The solid line represents the absorbance at 280 nm of the eluent. The inset shows the SDS-polyacrylamide gel electrophoresis analysis of aliquots from peaks 1a, 2a, and 3a (25 kDa, 13 kDa, and 11 kDa respectively, see text). **(B)** The pooled 25 kDa fractions were lyophilized, dissolved in 0.2% HOAc, and rechromatographed on a second C_{18} column. The absorbance of the effluent was monitored at 214 nm. The peak corresponding to the 25 kDa species was collected (peak 1b). The inset shows the SDS-polyacrylamide gel electrophoresis analysis of this pool. The molecular weight markers are the same as shown in Figure 1C.

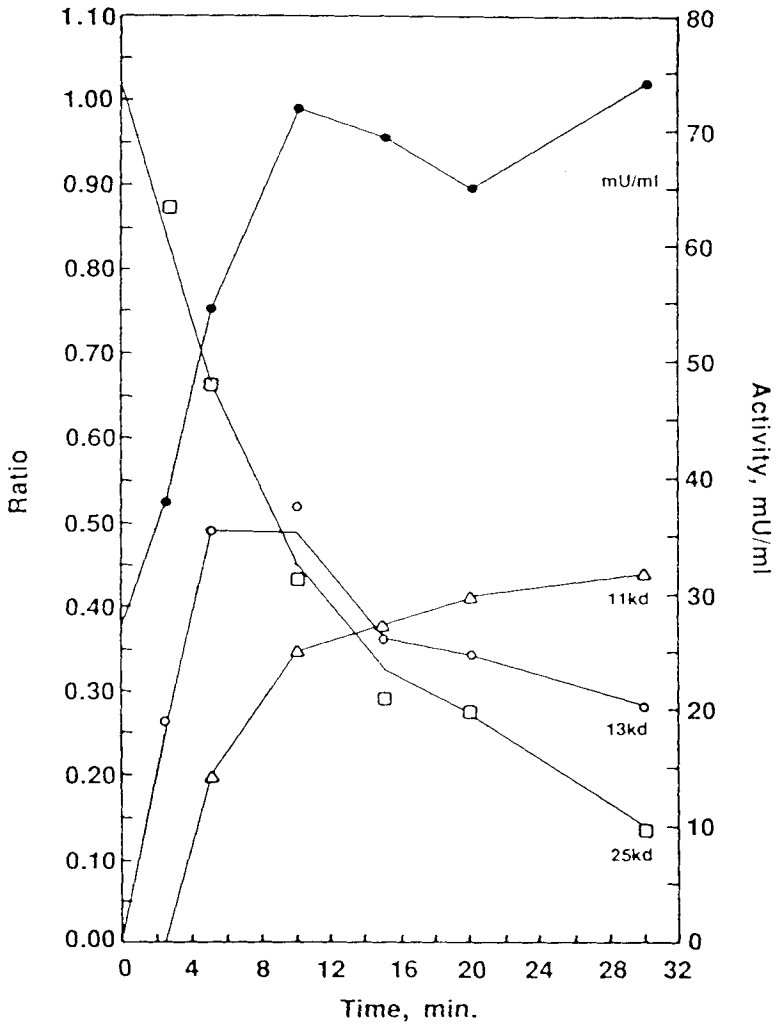

Fig 7. Time course of the in vitro autoprocessing of the 25kDa^{AR68} protease precursor. The 25 kDa precursor was redissolved in 0.2% HOAc at ~1 mg/ml and diluted to a final concentration of 25 μg/ml in 50 mM Tris-HCl, pH 7.0, 1 M urea, 250 mM NaCl, 0.01% Triton X-100, 1 mM DTT, and 0.5 mM EDTA. The diluted sample was incubated at 37°C. At 0, 2.5, 5, 10, 15, 20, and 30 min, aliquots were withdrawn for activity assay and for quantification of the residual 25 kDa precursor as well as of the newly formed 13 and 11 kDa protease species by RP-HPLC assay. The protein data are shown as the ratio of the amount present at a given time point to either the starting amount for the 25 kDa precursor (open squares) or the amount expected if the precursor were quantitatively converted to the 13 kDa intermediate (open circles) or the 11 kDa mature protease (open triangles). The closed circles show the increase in the peptidolytic activity.

preparations. It should be pointed out that the two enzyme assays are not strictly comparable; they used an octapeptide substrate while we used a nonapeptide substrate, they used 100 mM NaCl while we used 200 mM, and their assay was performed at pH 5.5 while we used pH 6.0. On the other hand, in our experience,[21,23] there is no significant difference in the kinetics of peptidolysis between octa- and nonapeptide substrates, between assays performed at pH 5.5 or 6.0, or between assays performed in 0.1 or 0.2 M NaCl. The difference in the construct used for expression is also not responsible for the lower activity of their enzyme. We have made a construct similar to theirs, expressed the protein in *E. coli*, purified the enzyme, and examined its kinetic parameters. Although the purification protocol for this enzyme is very different from the scheme we use for 11kDa^{AR58}, the K_m and k_{cat} values were essentially identical (unpublished results). It is possible that the protease encoded by the New York 5 HIV isolate they used may be less efficient than the protease encoded by the HIV BH10 isolate that we used.

Our earlier work[6] and the work presented here demonstrate that the HIV 1 protease is capable of catalyzing its own processing from a larger precursor in vivo. Mous et al.[15] have described a similar processing of HIV protease in vivo in *E. coli*. Their precursor is somewhat different from ours, but they also observed the appearance of a stable 11 kDa product in bacterial cells. In our system, we have

now shown that this processing occurs at the two sites in the *pol* reading frame that correspond to the consensus cleavage sequence, Ser/Thr-Xxx-Yyy-Tyr/Phe-Pro. In addition, we have also shown that the purified 25 kDa precursor, 25kDa^{AR68}, is capable of correctly autoprocessing in vitro. Indeed, the order of the processing steps can be determined from our data. Clearly, the NH_2-terminal extension is removed first, generating the 13 kDa intermediate. Accompanying this cleavage there is a 1.5- to 2-fold increase in the apparent activity of the preparation. This increase is maintained during the subsequent processing step that yields the mature COOH-terminus of the enzyme. We did not observe the appearance of significant levels of the potential 23 kDa intermediate that would form if processing of the COOH-terminal extension occurred before cleavage of the NH_2-terminal extension. Our data, however, cannot rule this out as a minor pathway. Both the 25 kDa precursor and the 13 kDa intermediate are active in the peptidolysis assay; although, the k_{cat} of the precursor is ~10-fold lower than that of the mature 11 kDa protease. From the proximity of the NH_2-terminus of the molecule to the active site in the crystal structure of the protease, Navia et al.[28] have suggested that the NH_2-terminal processing step could occur in *cis* while the COOH-terminal step would have to be performed in *trans*. Such a mechanism would be entirely consistent with our observations on the order of processing and the presence of an NH_2-terminal extension on the precursor could explain its lower k_{cat} through steric hindrance of the active site. It should be pointed out that the activities measured are under optimal in vitro conditions, i.e., high salt and pH 6.0, and toward a nonapeptide substrate. The quantitative relationship between this peptidolytic activity and the proteolytic activity of the enzyme toward its natural substrates, Pr55^{gag} and Pr160$^{gag-pol}$, is not clear; however, by inactivation of the precursor by chemical means or through active site modification, we will be able to directly compare the peptidolytic and proteolytic activity of the 11 kDa mature protease. This is an important control for design of active site inhibitors.

Finally, it is interesting to note that the enzyme requires one or both of the cysteine residues to be free in order to be active. The evidence for this is 3-fold: (1) In our previous work we have shown that the 11kDa^{AR58} protease can be inactivated by treatment with the sulfhydryl specific reagents, *N*-ethylmaleimide or iodoacetamide.[21] (2) In this paper, we demonstrate that the reversible oxidation of one or both of the cysteine residues can occur and that the oxidized protein is inactive. The activity can be restored following mild reduction. Furthermore, this oxidation is most likely the formation of an intramolecular disulfide bond, since such a bond would readily form between two cysteines, adjacent in space, under the storage conditions we employed. An intermolecular crosslink is ruled out by the observation of monomeric protease on nonreducing SDS-polyacrylamide gel electrophoresis and harsher oxidations, e.g., methionine to methionine sulfoxide or cysteine to cysteic acid, would not be as readily reversible as that observed. (3) The protease is also reversibly inactivated by sodium tetrathionate (data not shown), which is known to react with sulfhydryl groups. This inactivation can be reversed by treatment with DTT, again implicating a cysteine sulfhydryl. The enzyme is a dimeric, aspartyl-pro-tease[21,24,27,28]; hence, one or both of the cysteines must make important contributions to its overall structure and ultimately to the formation of the active site. From the three-dimensional structure,[28] it is hard to see how blockage of either cysteine by *N*-ethylmaleimide or tetrathionate could lead to steric hindrance of the active site. It is also hard to see how an intrachain disulfide could form between Cys-67 and Cys-95; however, our data show that sulfhydryl reagents do inactivate the enzyme and that it is possible for such a disulfide to occur under mild conditions. These data suggest that the structure of the protein is relatively dynamic under physiological conditions and the reactions which occur at the cysteines block the transition from an inactive to an active conformation. These apparently deleterious aspects of the protease could serve an important control function as it would be detrimental to the virus to have a highly active protease persist after the essential processing steps are performed. Thus, the dynamic structure and the readily oxidizable cysteines could serve as a built-in inactivation mechanism in order to spare the virus from harmful degradative activity.

In summary, we have expressed in bacteria, purified, and chemically characterized both the authentic, mature 11 kDa form of the HIV protease and a 25 kDa precursor having both NH_2- and COOH-terminal extensions. The 11 kDa form results from correct autoprocessing of the 25 kDa precursor in vivo. The purified 25 kDa precursor is also capable of correct autoprocessing in vitro. Both forms of the enzyme are active in a highly specific peptidolysis assay, with the mature form having a 10-fold higher k_{cat} than the precursor. The availability of both forms will allow us to study the effects of structural differences on the activities of the two enzymes.

ACKNOWLEDGMENTS

The authors wish to thank Marty Rosenberg, Irwin Chaiken, and Brian Metcalf for support, helpful discussions, and critical reading of the manuscript. This work was supported in part by Grant AI24845 from the National Institutes of Health.

NOTE ADDED IN PROOF

Recently, Wlodawer, et al.[39] reported the crystal structure of synthetic HIV-1 protease. Unlike the structure reported by Navia, et al.,[28] Wlodawer's group were able to determine the position of the NH_2-terminal 5 amino acids. These amino acids form a β-strand that interdigitates with the COOH-terminal β-strand of the other subunit in the homodimer to form part of the dimer interface. Thus, it is unlikely that intramolecular processing of the NH_2-terminus, as suggested by Navia, et al., can occur. Furthermore, since Cys-95 is found in this interfacial region, it is likely that blocking Cys-95 with iodoacetamide, N-ethylmaleimide, or tetrathionate as we describe, does not inactivate the enzyme by steric hindrance of the active site, but rather by destabilization of the dimer.

REFERENCES

1. Ratner, L., Haseltine, W., Patarca, R., Livak, K., Starcich, B., Josephs, S., Doran, E., Rafalski, J., Whitehorn, E., Baumeister, K., Ivanoff, L., Petteway, S., Pearson, M., Lautenberger, J., Papas, T., Ghrayeb, J., Chang, N., Gallow, R., Wong-Staal, F. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature (London) 313:277–284, 1985.
2. Sanchez-Pescador, R., Power, M.D., Barr, P.J., Steimer, K.S., Stempien, M.M., Brown-Shimer, S.L., Gee, W.W., Renard, A., Randolph, A., Levy, J.A., Dina, D., Luciw, P.A. Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). Science 227:484–492, 1985.
3. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., Alizon, M. Nucleotide sequence of the AIDS virus. Cell 40:9–17, 1985.
4. Muesing, M.A., Smith, D.H., Cabradilla, C.D., Benton, C.V., Lasky, L.A., Capon, D.J. Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. Nature (London) 313:450–458, 1985.
5. Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A.F., Scolnick, E.M., Sigal, I.S. Active human immunodeficiency virus protease is required for viral infectivity. Proc. Natl. Acad. Sci. U.S.A. 85:4686–4690, 1988.
6. Debouck, C., Gorniak, J.G., Strickler, J.E., Meek, T.D., Metcalf, B.W., Rosenberg, M. Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the *gag* precursor. Proc. Natl. Acad. Sci. U.S.A. 84:8903–8906, 1987.
7. Graves, M.C., Lim, J.J., Heimer, E.P., Kramer, A. An 11-kDa form of human immunodeficiency virus protease expressed in *Escherichia coli* is sufficient for enzymatic activity. Proc. Natl. Acad. Sci. U.S.A. 85:2449–2453, 1988.
8. Copeland, T.D., Oroszlan, S. Genetic locus, primary structure, and chemical synthesis of human immunodeficiency virus protease. Gene Anal. Techn. 5:109–115, 1988.
9. Lillehoj, E.P., Salazar, R.H.R., Mervis, R.J., Raum, M.G., Chan, H.W., Ahmad, N., Venkatesan, S. Purification and structural characterization of the putative *gag-pol* protease of human immunodeficiency virus. J. Virol. 62:3053–3058, 1988.
10. Pearl, L.H., Taylor, W.R. Sequence specificity of retroviral proteases. Nature (London) 328:482, 1987.
11. Lightfoote, M.M., Coligan J.E., Folks, T.M., Fauci, A.S., Martin, M.A. Venkatesan, S. Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus. J. Virol. 60:771–775, 1986.
12. di Marzo Veronese, F., Copeland, T., Vico, A., Rahman, R., Oroszlan, S., Gallo, R.C., Sarngadharan, M.G. Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. Science 231:1289–1291, 1986.
13. Farmerie, W.C., Loeb, D.D., Casavant, N.C., Hutchison, C.A., III, Edgell, M.H., Swanstrom, R. Expression and processing of the AIDS virus reverse transcriptase in *Escherichia coli*. Science 236:305–308, 1987.
14. Hansen, J., Billich, S., Schulze, T., Sukrow, S., Moelling, K. Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. EMBO J. 7:1785–1791, 1988.
15. Mous, J., Heimer, E.P., Le Grice, S.F.J. Processing protease and reverse transcriptase from human immunodeficiency virus type I polyprotein in *Escherichia coli*. J. Virol. 62:1433–1436, 1988.
16. Giam, C.-Z., Boros, I. *In vivo* and *in vitro* autoprocessing of human immunodeficiency virus protease expressed in *Escherichia coli*. J. Biol. Chem. 263:14617–14620, 1988.
17. Nutt, R.F., Brady, S.F., Darke, P.L., Ciccarone, T.M., Colton, C.D., Nutt, E.M., Rodkey, J.A., Bennet, C.D., Waxman, L.H., Sigal, I.S., Anderson, P.S., Veber, D.F. Chemical synthesis and enzymatic activity of a 99-residue peptide with a sequence proposed for the human immunodeficiency virus protease. Proc. Natl. Acad. Sci. U.S.A. 85: 7129–7133, 1988.
18. Schneider, J., Kent, S.B.H. Enzymatic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease. Cell 54:363–368, 1988.
19. Darke, P.L., Nutt, R.F., Brady, S.F., Garsky, V.M., Ciccarone, T.M., Leu, C.-T., Lumma, P.K., Freidinger, R.M., Veber, D.F., Sigal, I.S. HIV-1 protease specificity of peptide cleavage is sufficient for processing of *gag* and *pol* polyproteins. Biochem. Biophys. Res. Commun. 156:297–303, 1988.
20. Darke, P.L., Leu, C.-H., Davis, L.J., Heimbach, J.C., Diehl, R.E., Hill, W.S., Dixon, R.A.F., Sigal, I.S. Human immunodeficiency virus protease. Bacterial expression and characterization of the purified aspartic protease. J. Biol. Chem. 264:2307–2312, 1989.
21. Meek, T.D., Dayton, B.D., Metcalf, B.W., Dreyer, G.B., Strickler, J.E., Gorniak, J.G., Rosenberg, M., Moore, M.L., Magaard, V.M., Debouck, C. Human immunodeficiency virus-1 (HIV-1) protease expressed in *E. coli* behaves as a dimeric aspartyl protease. Proc. Natl. Acad. Sci. U.S.A. 86:1841–1845, 1989.
22. Billich, S., Knoop, M.-T., Hansen, J., Strop, P., Sedlacek, J., Mertz, R., Moelling, K. Synthetic peptides as substrates and inhibitors of human immune deficiency virus-1 protease. J. Biol. Chem. 263:17905–17908, 1988.
23. Moore, M.L., Bryan, W.M., Fakhoury, S.A., Magaard, V.W., Huffman, W.F., Dayton, B.D., Meek, T.D., Hyland, L., Metcalf, B.W., Strickler, J.E., Gorniak, J.G., Debouck, C. Substrates and inhibitors of the HIV-1 protease. Biochem. Biophys. Res. Commun. 159:420–425, 1989.
24. Katoh, I., Yasunaga, T., Ikawa, Y., Yoshinaka, Y. Inhibition of retroviral protease activity by an aspartyl proteinase inhibitor. Nature (London) 329:654–656, 1987.
25. Seelmeier, S., Schmidt, H., Turk, V., von der Helm, K. Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. Proc. Natl. Acad. Sci. U.S.A. 85:6612–6616, 1988.
26. Loeb, D.D., Hutchinson, C.A., III, Edgell, M.H., Farmerie, W.G., Swanstrom, R. Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases. J. Virol. 63:111–121, 1988.
27. Pearl, L.H., Taylor, W.R. A structural model for the retroviral proteases. Nature (London) 329:351–354, 1987.
28. Navia, M.A., Fitzgerald, P.M.D., McKeever, B.M., Leu, C.-T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L., Springer, J.P. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature (London) 337:615–620, 1989.
29. Devare, S.G., Shatzman, A., Robbins, K.C., Rosenberg, M., Aaronson, S.A. Expression of the PDGF-related transforming protein of simian sarcoma virus in *E. coli*. Cell 36: 43–49, 1984.
30. Watt, R.A., Shatzman, A.R., Rosenberg, M. Expression and characterization of the human c-*myc* DNA-binding protein. Mol. Cell. Biol. 5:448–456, 1985.
31. Baker, T.A., Grossman, A.D., Gross, C.A. A gene regulating the heat shock response in *Escherichia coli* also affects proteolysis. Proc. Natl. Acad. Sci. U.S.A. 81:6779–6783, 1984.

32. Rosenberg, M., Ho, Y.-S., Shatzman, A. The use of pKC30 and its derivatives for controlled expression of genes. Meth. Enzymol. 101:123–138, 1983.
33. Read, S.M., Northcote, D.H. Minimization of variation in the response to different proteins of the Coomassie Blue G dye binding assay for protein. Anal. Biochem. 116:53–64, 1981.
34. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248–254, 1976.
35. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680–685, 1970.
36. Merril, C.R., Goldman, D., Sedman, S.A., Ebert, M.H. Ultrasensitive stain for proteins in polyacrylamide gels show regional variation in cerebrospinal fluid proteins. Science 211:1437–1438, 1981.
37. Jones, B.N., Gilligan, J.P. *o*-Phthaldialdehyde precolumn derivatization and reversed-phase high-performance liquid chromatography of polypeptide hydrolysates and physiological fluids. J. Chromatograph. 266:471–482, 1983.
38. Shaw, G.M., Hahn, B.H., Arya, S.K., Groopman, J.E., Gallo, R.C., Wong-Staal, F. Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired imune deficiency syndrome. Science 226:1165–1171, 1984.
39. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., Kent, S.B.H. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science 245:616–621, 1989.