Autoprocessing of the human immunodeficiency virus type 1 protease precursor expressed in Escherichia coli from a synthetic gene

Viviane Valverde, Pierre Lemay, Jean-Michel Masson, Bernard Gay and Pierre Boulanger

1Centre de Transfert en Biotechnologie-Microbiologie, CNRS-UA544, INSA, Complexe Scientifique de Rangueil, 31077 Toulouse and 2Laboratoire de Virologie et Pathogénèse Moléculaires, Faculté de Médecine, 34060 Montpellier, France

A gene encoding an N-terminally extended precursor of 107 residues of the human immunodeficiency virus type 1 protease (PR107) was chemically synthesized and cloned into a bacterial expression vector, under the control of the araB promoter. PR107 was expressed alone or fused in phase to the amino or carboxy terminus of the bacterial β-galactosidase (β-gal). The yield of protease and β-gal was found to be significantly higher when the gene for PR107 was cloned upstream of the Escherichia coli lacZ gene (PR107–β-gal). Comparisons of the level of cloned protein expression between protease precursor and mature form suggested that this enhanced expression was due to the additional 5' sequence of the PR107 gene, and occurred at the post-transcriptional level. Autoprocessing of protease precursor and its release from the β-gal fusion protein were analysed using wild-type and mutated cleavage sites. Mutations were introduced at amino acids downstream of the F–P scissile bond, at positions P4' and P5' in the C-terminal site (TLNF*PISP), and at position P3' in a consensus N-terminal site (TLNF*PQITL) placed at the protease–β-gal junction. The data obtained suggested that (i) autoprocessing at the carboxy-terminal F–P bond was not significantly influenced by the presence of the N-terminal precursor sequence, (ii) P4' and P5' substitutions in the C-terminal site had no effect on cleavage, and (iii) P3' in the N-terminal site tolerated a wide variety of substitutions.

Introduction

Specific processing of the human immunodeficiency virus type 1 (HIV-1) gag-pol polyprotein by the virus-encoded protease yields the structural gag gene products p24CA, p17MA and p15NC, as well as non-structural proteins, reverse transcriptase, endonuclease and protease (reviewed in Cann & Karn, 1989; Wills & Craven, 1991). The HIV-1 protease originates from a large gag–prt–pol polyprotein precursor, presumably as a result of intermolecular autoprocessing events (Lillegoj et al., 1988; Mous et al., 1988). It belongs to the aspartyl protease family and its active site contains the consensus D-T-G sequence (reviewed in Krausslich & Wimmer, 1988). It is active as a dimer (Katoh et al., 1989; Miller et al., 1989b) and is inhibited in vitro by pepstatin (Seelmeier et al., 1988; Katoh et al., 1987). Since specific proteolytic cleavages are essential for assembly of infectious HIV virions (Kohl et al., 1988; Peng et al., 1989; Gelderblom, 1991), the protease represents one of the possible targets for enzyme-directed anti-AIDS therapy (reviewed in Skalka, 1989; Tomasselli et al., 1991). HIV-1 protease has been produced in bacteria (Debouck et al., 1987; Graves et al., 1988), chemically synthesized (Schneider & Kent, 1988), crystallized (Miller et al., 1989a, 1990; Navia et al., 1989) and co-crystallized with peptide-based specific inhibitors (Erickson et al., 1990; Miller et al., 1989b; Wlodawer et al., 1989).

The aim of the present study was to analyse the mechanism of autocatalytic processing of HIV-1 protease at its N and C termini, and its subsequent release from a fusion protein. For this purpose, as an alternative to in vitro chemical or in vivo biological synthesis, the HIV protease was expressed from a synthetic gene cloned into a bacterial expression vector. We cloned the prt gene in Escherichia coli, unfused or fused in phase to the 5' or 3' end of the bacterial lacZ gene (upstream or downstream position, respectively). Our approach has several advantages compared to previously described production systems. (i) A synthetic sequence has more versatility for further genetic manipulations; (ii) the protease gene sequence was designed for optimal codon usage in E. coli; (iii) since the cloned protease was expected to be toxic for the recipient cell (Hostomsky et al., 1989), the protease gene was cloned under the strong but tightly regulated araB promoter (Cagnon et al., 1991); (iv) the yield of the β-galactosidase (β-gal)-fused gene products could be monitored easily using a simple β-gal enzymic assay; (v) purification of HIV-1 protease.
might be achieved by affinity chromatography of protease-β-gal fusion on immobilized β-gal substrate or anti-β-gal antibody. Bacterial β-gal has also been recently used to monitor the HIV protease activity upon insertion of one of its specific sites into the lacZ gene (Baum et al., 1990).

The HIV-1 protease sequence is flanked by two processing sites, SFNF*PQIT at its N terminus and TLNF*PISPI at its C terminus. Cleavage of the protease precursor and release from the protease-β-gal fusion protein was analysed by mutagenesis of amino acids downstream of the F-P scissile bond, at positions P4' and P5' in the C-terminal processing site and at position P3' in a consensus N-terminal site placed at the protease-β-gal junction. The tolerance to amino acid residue substitutions was thus evaluated from the rate of protease-β-gal cleavage. We found that mutations at subsites P4' and P5' had no apparent effect on cleavage, but that the P3' position in the N-terminal site showed some sensitivity to amino acid substitutions with strong effects on secondary structure.

Methods

Bacterial strains and plasmids. E. coli strain TG-1 (Amersham) was used for cloning, and MC1061 (Casadaban & Cohen, 1980) for gene expression. Cells were grown in LB medium supplemented with ampicillin (100 μg/ml; LBamp). The unfused protease gene was constructed from oligonucleotide blocks directly assembled into pCris10, a derivative of pKK233-2 (Amman & Brosius, 1985) which differs from pKK233-2 by its synthetic cloning cassette (V. Valverde et al., unpublished results). β-gal–protease fusion constructs were directly assembled into pARA14. Plasmid pARA14 contains an arabinose-inducible promoter, termed araB (Cagnon et al., 1991), and constituted our expression vector. For induction of gene expression, overnight cultures were diluted 20-fold in LBamp, and incubated at 37°C until they reached an optical density at 600 nm (OD600) of 0.5. Arabinose was then added to the cultures to a final concentration of 0.2%.

Nomenclature. PR100 and PR107 refer to the mature protease and protease precurser of 100 and 107 amino acid residues, respectively, including the initiator codon, methionine. When the prt gene was fused to the 5' end of the lacZ gene (upstream position), the resulting in-phase fusion protein was called PR107-β-gal or PR100-β-gal. When prt was fused to the 3' end of lacZ (downstream position) the name was β-gal–PR107. No β-gal-PR100 fusion protein was constructed since this protein would have lacked the protease N-terminal processing site at the β-gal–protease junction. The protease mutant named G33 had a dipeptide Arg-Arg at positions P4' and P5' from the scissile bond at the PR107-β-gal junction was substituted for the neutral Pro-Gly processing site, but containing a suppressible amber stop codon in lieu of the isoleucine codon at position P3', was introduced at the downstream PR107-β-gal junction, the resulting plasmid was referred to as PRamb5'LZ. Amber suppressions at position P3' in this junction sequence gave rise to P3' mutants. A schematic drawing of our different gene constructs is presented in Fig. 1(e).

Gene constructions. Oligonucleotides were synthesized on an Applied Biosystems 380A DNA synthesizer (Applied Biosystems) and purified as trityl derivatives on Nensorb Prep columns (NEN Research products) according to the manufacturer's specifications. They were enzymatically phosphorylated as previously described (Normanly et al., 1986). Two different HIV-1 prt genes were constructed. The first one encoded the mature form of protease of 100 amino acid residues, starting its sequence with the Met(1)–Pro(2) dipeptide whereas the second encoded a precursor form of 107 amino acids. The PR107 gene contained an additional 5' nucleotide sequence encoding the heptapeptide (M)GTVSFNF (not counting the initiator methionine) present within the gag–prt–pol polyprotein, at the gag–protease junction, thus reconstituting the original Phe–Pro dipeptide of the N-terminal protease processing site (Darke et al., 1988, 1989).

Both genes for the unfused protease constructs PR100 and PR107 were assembled stepwise from 20 complementary synthetic oligonucleotides annealed and ligated into three adapting building blocks (Fig. 1b). Oligonucleotides corresponding to the protease coding strand were given odd numbers, and the even numbers corresponded to the non-coding strand (Fig. 1c). Step 1, oligonucleotides C1 to C8 were annealed and cloned between the HindIII and BglII restriction sites of pCris10. Step 2, the central part of the gene (oligonucleotides M1 to M6) was inserted between the BglII and NalI sites. Step 3, the 5' end of the PR100-encoding gene was reconstituted from oligonucleotide pairs N1-N2, N5-N6 and N7-N8, cloned between NalI and NcoI; for the 5' end of PR107, N5-N4 was inserted in lieu of N1-N2 (Fig. 1c, and inset e). For optimal expression, the sequence was designed according to the codon usage of E. coli (Fig. 1c). All gene constructs were verified by DNA sequencing (Sanger et al., 1977). PR100 and PR107 genes were then inserted into the pARA14 expression vector, under the control of the arabinose-inducible promoter araB (Cagnon et al., 1991).

In-phase β-gal fusions were directly constructed into the pARA14 expression vector already containing the gene for PR100 or PR107. The downstream fusion β-gal–PR107 was obtained by insertion of the plZ4 Neol β-gal gene cassette (V. Valverde et al., unpublished) into the NcoI site of PR107. The two upstream fusions PR100-β-gal and PR107-β-gal were generated by ligating the 5' end of the lacZ gene to the 3' end of the PR100 or PR107 gene. This was achieved by first cloning the L1 linker sequence (Fig. 1d) at the 3' end of the prt gene. The plZ2 cassette (V. Valverde et al., unpublished) was then inserted between the SalI and HindIII sites, within the L1 and ARA14 sequences, respectively. The L1 linker was then mutated into the L0 linker (Fig. 1d), to obtain the PR100-β-gal and PR107-β-gal fusions.

Site-directed mutagenesis at the catalytic site and protease-β-gal junction. The phosphorothioate method (Taylor et al., 1985) was used to introduce two substitutions into the active site of the protease, G and T substituting for A and C at nucleotides 100 and 101 of the PR107 gene, respectively (Fig. 1d). This created a new KpnI site and changed the D (33) residue of the D-T-G triad (position 25 in the mature enzyme) into a glycine. This gave rise to the unfused G33 and β-gal-fused G33–β-gal mutant constructs (Fig. 1a).

The protease-β-gal junction in PR100-β-gal and PR107-β-gal fusions consisted of the sequence TLNF*PISPI, representing the C-terminal processing site of the protease. By our cloning strategy (see above), the RR-β-gal fusion was obtained first, since it contained the L1 linker encoding the junction sequence LNFPISPRRG. Using the same method (Taylor et al., 1985), the L1 linker was then mutated into the L0 sequence (Fig. 1d), encoding the sequence LNFPISPGI. This latter sequence better mimicked the natural C-terminal processing site LNFPISPIG1, in terms of side chain electric charges and overall conformation.

The L1 linker was then mutated into the L0 linker (Fig. 1d), to obtain the PR100-β-gal and PR107-β-gal fusions.
Amber mutation suppressor tRNA scanning. A double-stranded oligonucleotide sequence (coding strand: 5' TGCAGCTCT-GAACCCTCTCTAGACTGGGGATCCCA 3') encoding the consensus N-terminal processing site TLNF*PQITL was synthesized, with an amber codon replacing the isoleucine codon at position P3' of the scissile bond. The TLNF*PQITL-encoding nucleotide sequence was then inserted at the PR107 fl-gal junction, between the PRamb3'LZ and the fl-gal gene. 

Fig. 1. (a) Schematic representation and nomenclature of the various protease gene (pr) constructs with the N- and C-terminal autoprocessing sites. The symbols and abbreviations used are: β-gal, β-galactosidase; araB, arabinose-inducible promoter; OR1, ORF-1 promoter; TE1, transcription terminator; Nt, amber mutation; G, D to G substitution in the protease active site; X, amino acid specified by amber suppressor tRNA at the P3' position in the downstream processing site consensus to the N-terminal site. (b) Diagram of the stepwise construction of HIV-1 mature protease gene (PR100). Three successive building blocks with appropriate sticky ends were inserted into the pCris10 plasmid (Cagnon et al., 1991). (c) Nucleotide sequence of the synthetic gene for the protease precursor of 107 residues (PR107), coding for the protease of 99 amino acids with an initiation methionine residue (PR100), and an additional N-terminal heptapeptide proline-isoleucine dipeptide at the P4' and P5'' subsites was replaced by Arg–Arg. In the inset (e) are shown the two complementary oligonucleotides used to construct the gene for PR100 protease, N1 and N2 replacing N3 and N4 at the 5' end.

Autoprocessing of HIV-1 protease 641

(c) Synthetic gene sequences

| N1 | CATGCTTACGCTGACCGGCTGAGG
| N2 | CATGCTTACGCTGACCGGCTGAGG

(d) Junction sequences

L0 linker: CTGACCTCTCTCTTGGGACCTATTTTTTT
LeuAmpFmProlIederFmProlIey1

L1 linker: CTGACCTCTCTCTTGGGACCTATTTTTTT
LeuAmpFmProlIederFmProlIey1

(e) Junction sequences

PRamb3' fl-gal

N3-N8 (105 bp) M1-M6 (102 bp) C1-C8 (120 bp)

NcoI [NcoI] NarI EcoRV BglII Kpnl HindIII

N3-N8 (105 bp) M1-M6 (102 bp) C1-C8 (120 bp)

NcoI NarI BglII NarI HindIII
protease-β-gal fusion protein and β-gal released from the protease, respectively. The 127K/116K doublets were scanned on anti-β-gal immunoblot patterns of SDS-polyacrylamide gels.

**Enzyme assays.** β-Gal activity was assayed in extracts from toluene-permeabilized cells by hydrolysis of o-nitrophenyl-β-D-galactoside (ONPG) (Miller, 1972). For protease assay, 1 ml aliquots of arabinose-induced cell culture taken at different time intervals were pelleted, washed in TN buffer (20 mM-Tris-HCl pH 7.5, 0.1 M-NaCl) and lysed in 50 μl of lysis buffer at 4 °C overnight with gentle mixing. Lysis buffer was made of 30 mM-Tris–HCl pH 7.5, 1 mM-DTT, 0.01% lysozyme, 0.1% NP40, 1 mg/ml DNase I and a cocktail of protease inhibitors (Darke et al., 1988). The precursor to p24CAw produced in a baculovirus expression system (Royer et al., 1991) was used as the substrate for protease. Aliquots (10 μl) of gag polyprotein (0.5 to 1.0 μg) in 0.2 M-sodium phosphate buffer pH 6.5, 1 M-NaCl, 5% glycerol and 0.25% NP40 were incubated with 10 μl of bacterial cell lysate at 27 °C and the reaction was stopped at different time intervals by addition of 20 μl of SDS-PAGE sample buffer. The gag precursor cleavage products were analyzed by SDS-PAGE and immunoblotting with gag-specific antibody.

**Biological and immunological analyses.** Bacterial proteins were analysed by SDS-PAGE (15%) in a discontinuous buffer system (Laemmli, 1970). In a typical experiment, 1 ml bacterial cell culture was pelleted, washed with TN buffer, resuspended in 100 μl of TN buffer, mixed with 100 μl of SDS–urea–2-mercaptoethanol sample buffer and denatured at 100 °C for 2 min. Gels were stained or electrically transferred to nitrocellulose membrane (BA85, Schleicher & Schüll) for 45 min at 180 mA in a semi-dry apparatus (Millipore SDE). Membranes were reacted with anti-β-gal (Research Plus) or anti-HIV-1 protease rabbit polyclonal antibody (a gift from Dr E. Cheng, DuPont) and a phosphatase-labelled anti-ribonuclease conjugate (Sigma). Gag polyprotein products were detected on blots by reaction with an anti-pr55-p24CAw mouse monoclonal antibody (Epitope Inc.; Epitope Inc.) and a phosphatase-labelled anti-mouse IgG conjugate (Sigma). ELISA was performed using commercially available bacterial β-gal (Sigma) and anti-β-gal antibody (Research Plus).

For protein microsequencing, proteins were partially purified by acetone fractionation from bacterial cell lysates (Hansen et al., 1988). The acetone precipitate was then chromatographed on FPLC-MonoS column in 50 mM-sodium 2-(N-morpholino)ethanesulphonic acid buffer pH 6.5, 1 mM-disodium EDTA, 10% glycerol (Pharmacia). Protease was eluted at 0.2 M-NaCl and its purification was achieved by preparative SDS-PAGE in a Tricine-buffered system (Schägger & von Jagow, 1987). For β-gal and protease, the 116K and 11K bands were transferred to Immobilon membranes (Ploug et al., 1989) and amino-terminal sequences were determined using an Applied BioSystems 470A protein sequencer coupled to an Applied BioSystems 120A PTH Analyser.

**Electron and immunoelectron microscopy.** E. coli cells fixed in 2.5% glutaraldehyde in 0.1 M-phosphate buffer pH 7.5, for 1 h at 4 °C, were post-fixed with 2% osmium tetroxide in water and embedded in Epon (Epox-812; Fullam). Immunogold staining (IGS) was carried out by successive reactions of cell sections with primary anti-β-gal or anti-protease rabbit antibody overnight at 4 °C (at a dilution of 10 μg/ml in Tris-buffered saline), and 5 nm colloidal gold-labelled anti-ribonuclease IgG antibody for 1 h at room temperature (single IGS reaction). For double IGS, specimens were first simultaneously incubated with mouse monoclonal anti-β-gal antibody (Sigma) and rabbit polyclonal anti-protease antibody, then with 1 nm gold-labelled anti-ribonuclease IgG antibody and 5 nm gold-labelled anti-mouse IgG antibody (EM-GAM1 and EM-GAMS; BioCell Research Lab). Specimens were post-stained with 0.5% uranyl acetate and examined under the Philips EM-300 electron microscope.

**Results**

**Expression of β-gal-fused and unfused HIV-1 protease and protease precursor in E. coli.**

The level of expression of protease and β-gal synthesized in bacterial cells under the control of the araB promoter was estimated by SDS–PAGE and immunoblotting, and quantified by the β-gal assay. In *E. coli* expressing the unfused construct PR107, a band at 11K reacting with anti-protease antibody appeared after 1 h of arabinose induction (Fig. 2, lane 6) and progressively increased with the time of induction (Fig. 2, lanes 7 and 8). A similar pattern was obtained with the G33 mutant (Fig. 2, lanes 9 to 12), whereas no specific 11K band was detected in the PR100 pattern (Fig. 2, lanes 1 to 4). With the fused construct PR107-β-gal, the anti-protease serum revealed a discrete band of 127K fusion protein as early as after 0.5 to 1 h of induction (Fig. 3a, lanes 5 and 6) and an intense band at 11K became visible after 2 h (Fig. 3a, lanes 7 and 8). Several other discrete bands of intermediate cleavage products, ranging from 80K to 20K in apparent *M*, were also seen on the blot (Fig. 3a, lanes 5 to 8). With the β-gal-fused mutant G33-β-gal, the 127K fusion band increased in intensity during the induction period, without detectable release of the 11K protease (Fig. 3a, lanes 13 to 16). In contrast to PR107, PR107-β-gal, G33 and G33-β-gal, a low level of expression of protease and fused β-gal–protease was obtained with PR100 and PR100-β-gal (Fig. 2, lanes 1 to 4; Fig. 4a, lanes 9 to 12), as well as with the downstream construct β-gal–PR107 (Fig. 3a, lanes 9 to 12). Immunoblotting with anti-β-gal antibody (Fig. 3b) confirmed the results obtained with the anti-protease serum (Fig. 3a). β-Gal was expressed at much higher levels with PR107-β-gal, G33-β-gal and RR-β-gal than with the other fusions β-gal–PR107 and PR100-β-gal (Fig. 3b and 4b).

The results of immunoblots were confirmed by β-gal assays performed on cell extracts: the level of β-gal expression with upstream fusions PR107-β-gal, G33-β-gal and RR-β-gal was ninfold higher than with the downstream fusion β-gal–PR107, and 15-fold higher than with PR100-β-gal (Table 1). The common feature between all the highly expressed genes was the presence of the additional S’ sequence encoding the N-terminal precursor protease heptapeptide. To determine whether this enhanced expression of the cloned genes occurred at the transcriptional or post-transcriptional level, the overall quantity of protease mRNA expressed upon arabinose induction by the different fusion constructs was estimated by slot-blot hybridization, using 32P-kinase-labelled oligonucleotide M6 as the ssDNA probe (Fig. 1c). No significant difference was found in the
Autoprocessing of HIV-1 protease

643

**Table 1. Enzymic assay of β-gal-fused protease yields by E. coli cells harbouring various expression vectors**

<table>
<thead>
<tr>
<th>Vector</th>
<th>β-Gal activity (U/ml)</th>
<th>Ratio 1†</th>
<th>Ratio2‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>pARA1-lacZ</td>
<td>610</td>
<td>1.0</td>
<td>—</td>
</tr>
<tr>
<td>β-gal-PR107</td>
<td>595</td>
<td>0.9</td>
<td>—</td>
</tr>
<tr>
<td>PR100-β-gal</td>
<td>313</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>PR107-β-gal</td>
<td>4850</td>
<td>7.9</td>
<td>15.5</td>
</tr>
<tr>
<td>G33-β-gal</td>
<td>3110</td>
<td>8.4</td>
<td>16.3</td>
</tr>
<tr>
<td>RR-β-gal</td>
<td>4444</td>
<td>7.3</td>
<td>14.2</td>
</tr>
</tbody>
</table>

† Ratio of protease-fused β-gal to pARA1-lacZ value.
‡ Ratio of protease-fused β-gal to PR100-β-gal value.

Fig. 2. SDS-PAGE and immunoblotting analysis of unfused HIV-1 protease expressed in *E. coli* under the control of araB promoter. Cells expressing PR100 (lanes 1 to 4), PR107 (5 to 8) and G33 mutant (9 to 12) were harvested after 0.5, 1, 2 and 4 h of arabinose induction, respectively. Lane M, prestained marker lane (18K and 14K markers are shown). Blot was reacted with anti-HIV-1 protease rabbit serum and phosphatase-labelled secondary antibody.

Table 1. Enzymic assay of β-gal-fused protease yields by *E. coli* cells harbouring various expression vectors*

<table>
<thead>
<tr>
<th>Vector</th>
<th>β-Gal activity (U/ml)</th>
<th>Ratio 1†</th>
<th>Ratio2‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>pARA1-lacZ</td>
<td>610</td>
<td>1.0</td>
<td>—</td>
</tr>
<tr>
<td>β-gal-PR107</td>
<td>595</td>
<td>0.9</td>
<td>—</td>
</tr>
<tr>
<td>PR100-β-gal</td>
<td>313</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>PR107-β-gal</td>
<td>4850</td>
<td>7.9</td>
<td>15.5</td>
</tr>
<tr>
<td>G33-β-gal</td>
<td>3110</td>
<td>8.4</td>
<td>16.3</td>
</tr>
<tr>
<td>RR-β-gal</td>
<td>4444</td>
<td>7.3</td>
<td>14.2</td>
</tr>
</tbody>
</table>

† Ratio of protease-fused β-gal to pARA1-lacZ value.
‡ Ratio of protease-fused β-gal to PR100-β-gal value.

* Figures in the table are values of β-gal activity, determined by hydrolysis of ONPG and expressed in units per ml (U/ml) of permeabilized cell extracts (Miller, 1972). Each value represents the average of five individual results obtained in triplicate experiments (S.E.M., 5 to 10% of the activity values). The β-gal-specific activity was 1 \times 10^5 U/mg, as estimated from stained gel scannings and ELISA. The value of 4444 to 4850 U/ml thus corresponded to 44 to 48 mg/l.

amount of protease mRNA expressed by the various constructs (data not shown), suggesting that the enhancing effect occurred at the post-transcriptional level.

The yield of HIV-1 protease obtained with the vector expressing the fusion construct PR107-β-gal was estimated by three indirect methods, all of them based on β-gal determination: (i) β-gal enzymic assays (Miller, 1972), from calculated specific activity of 10^5 units per mg of β-gal protein; (ii) ELISA, using commercial β-gal as the standard and anti-β-gal serum; (iii) Coomassie blue staining of the 127K/116K doublet band of protease-fused and released β-gal in SDS-PAGE, after calibrating the gel with purified β-gal. The protease (11K) represents about one-tenth of the mass of the fusion protein, in terms of polypeptide mass ratio (116K for β-gal). Thirty to 50 mg of PR107-β-gal fusion protein was obtained per litre of cell culture after 4 h of arabinose induction. Assuming a 100% efficiency for protease self-processing and release from the fusion protein, its theoretical recovery would be 3 to 5 mg of protease per litre, i.e. 1-2 to 2-0 mg per g of *E. coli* wet weight. A similar protease recovery (1 mg/l) was obtained from *E. coli* expressing a maltose-binding protein fusion construct (Louis et al., 1991).

Activity, cytotoxicity and cellular distribution of β-gal-fused and unfused protease and protease precursor in *E. coli*

The results of the immunoblot analysis, showing evidence of HIV-1 protease release from the wild-type β-gal fusion constructs and accumulation of non-cleaved fusion protein with the G33-β-gal mutant (Fig. 3 and 4), suggested that the cloned protease was enzymically active and capable of self-processing in vitro. It could therefore be anticipated that protease synthesized in *E. coli* upon arabinose induction would also be active on its natural gag substrate. To verify its specificity of cleavage in vitro, our protease was incubated with baculovirus-expressed 41K gag polyprotein as a substrate (pr41\textsuperscript{agg}). Pr41\textsuperscript{agg} comprised the p17MA and p25CA domains (amino acids 1 to 375 of the gag sequence), with only two unique protease sites (Mervis et al., 1988); Tyr–Pro at the p17–p24 junction (amino acids 132 and 133), and Leu–Ala at the p24–p25 junction (positions 363 and 364), the Met–Met site at 377 and 378 being eliminated by carboxy truncation (Royer et al., 1991).

As expected, the active site mutants G33 and G33-β-gal showed no detectable proteolysis of pr41\textsuperscript{agg} (not shown), whereas all the other protease clones were found to cleave the gag polyprotein substrate at its two specific sites, yielding the characteristic anti-gag antibody-reacting p24–p25 doublet (Fig. 5). The highest activity was obtained with PR107 and PR107-β-gal, which converted almost all the gag precursor into p24–p25CA\textsuperscript{agg} after 30 min of incubation (Fig. 5, lanes 3 and 7). The lower proteolytic activity shown by the extracts of cells
Fig. 3. SDS–PAGE and immunoblotting analysis of β-gal-fused protease clones PR107–β-gal (lanes 1 to 4 and 5 to 8), β-gal–PR107 (9 to 12) and G33–β-gal mutant (13 to 16), expressed in E. coli in the presence (+) or the absence (−) of arabinose inducer. The same blot was successively reacted with (a) anti-protease antibody and phosphatase-labelled conjugate, and (b) anti-β-gal antibody and its corresponding phosphatase-labelled conjugate. Cells were harvested after 0.5, 1, 2 and 4 h of induction. M, prestained M₆, markers (BRL, high M₆ range). Protease–β-gal fusion protein migrates as a 127K species, free β-gal and free protease as 116K and 11K protein bands, respectively.
Autoprocessing of HIV-1 protease

(a) pARA-lacZ

<table>
<thead>
<tr>
<th></th>
<th>(-)</th>
<th>(+)</th>
<th>(+)</th>
<th>(+)</th>
<th>Ara</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>0-5</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>0-5</td>
</tr>
<tr>
<td>β-gal</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>0-5</td>
<td>1</td>
</tr>
</tbody>
</table>

(b) pARA-lacZ

<table>
<thead>
<tr>
<th></th>
<th>(-)</th>
<th>(+)</th>
<th>(+)</th>
<th>(+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>0-5</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>β-gal</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>0-5</td>
</tr>
</tbody>
</table>

Fig. 4. SDS-PAGE and immunoblotting analysis of control pARA-lacZ (lanes 1 to 8) and β-gal-fused protease clones PR100-β-gal (lanes 9 to 12) and RR-β-gal mutant (13 to 16), expressed in E. coli. As in Fig. 3, the blot was successively reacted with (a) anti-protease antibody and phosphatase-labelled conjugate, and (b) anti-β-gal antibody and its corresponding phosphatase-labelled conjugate. Cells were maintained for 0-5, 1, 2 and 4 h in the presence (+) or absence (-) of arabinose. M, prestained Mr markers; β-gal, bacterial β-galactosidase (116K) from a commercial source. Note that the central portion of the blot, of no informative value, is not presented in this figure.

Fig. 5. Activity of bacterially expressed HIV-1 protease on its natural substrate in vitro. Extracts from E. coli cells harvested after 4 h of arabinose induction were incubated at 27 °C with baculovirus-expressed gag polypeptide pr41, precursor to p24-p25CAgag. Samples were withdrawn at 5, 15, 30 and 60 min, and analysed in SDS-PAGE and immunoblotting with mouse monoclonal anti-gag precursor antibody (Epiclon 5001, Epitope) and phosphatase-labelled conjugate. Lane M, prestained Mr markers; lane 0, control gag polypeptide incubated for 60 min with extract from non-induced cells harbouring PR107.
expressing PR100–β-gal and β-gal–PR107 (Fig. 5, lanes 9 to 12 and 13 to 16) reflected the lower levels of protease production from these two constructs (Fig. 2 to 4).

However, the difference in fused and unfused protease expression observed between some of our clones could be due to cytotoxicity of the protease and a negative feedback effect on protein synthesis. In addition, c.p.e. has been frequently found to be related to the cellular distribution of cloned foreign proteins. Arabinose-induced cells were therefore analysed with respect to their growth rate, and the cellular localization of unfused and β-gal-fused protease was examined by immunoelectron microscopy. As shown in Fig. 6, no detectable cytotoxicity was observed with the fusion constructs, and expression of only unfused PR107 and G33 proteases resulted in a slight reduction of the growth rate after 4 h of induction. The similar patterns shown by the clones expressing the active PR107 protease and the inactive G33 mutant suggested that the observed c.p.e. was due to the cellular accumulation of a foreign protein rather than to proteolytic activity of the protease per se.

When lysates from cells expressing unfused protease or β-gal alone were fractionated, most of the β-gal and protease activities were recovered in the soluble supernatant (not shown). No inclusion body was observed under the electron microscope, and IGS showed that the cloned proteins (protease or β-gal) were evenly distributed within the E. coli cytoplasm (not shown). In contrast, amorphous inclusions were found in cells expressing the two fused constructs PR107–β-gal and G33–β-gal. These intracellular inclusions reacted with both anti-β-gal and anti-protease antibodies in single IGS reactions (Fig. 7a, b). In the PR107–β-gal-expressing clone, the gold grain pattern was found to be different with anti-β-gal and anti-protease antibodies: β-gal molecules were observed inside and outside the inclusions, whereas protease was exclusively localized within the inclusion bodies (compare Fig. 7a and b). Double IGS, using 1 and 5 nm
colloidal gold particles specific for rabbit (anti-protease) and mouse (anti-β-gal) antibodies, respectively, confirmed that the inclusions contained both β-gal and protease molecules (Fig. 7c), and suggested that protease remained associated with the inclusion and in the vicinity of β-gal molecules after proteolytic processing.

**Autoprocessing of HIV-1 protease: proteolytic cleavage at the protease-β-gal junction**

For PR107-β-gal, as well as PR100-β-gal, cleavage at the F-P bond situated in the natural C-terminal site TLNF*PISP placed at the protease-β-gal junction was required to yield the mature protease (Fig. 1a). As already shown in Fig. 3a, the first band to appear at 0.5 h of induction in the PR107-β-gal pattern was a 127K fusion protein detected with anti-protease antibody. A protease band at 11K occurred in significant amounts after 2 h of induction, concurrently with a doublet at 127K and 116K (Fig. 3b). The 116K species was detected only with anti-β-gal antibody (Fig. 3b). No 11K band was visible in the G33-β-gal mutant although large quantities of fusion protein accumulated at 127K (Fig. 3a, b, lanes 15 and 16). All these data suggested that the protease was released from the β-gal fusion protein by a mechanism of specific self-processing at the C-terminal site of the protease and not by proteolysis due to bacterial proteases. Although produced in much lower yields, the 127K fusion band of PR100-β-gal was cleaved at the same time of induction as that of PR107-β-gal (Fig. 4), with the occurrence of detectable protease activity in the cell lysates (Fig. 5, lanes 11 and 12). This result suggested that, in the protease-β-gal fusion construct, cleavage at the C-terminal F-P bond of protease was not significantly influenced by the presence of cleaved or uncleaved N-terminal precursor sequence.

**Specificity of cleavage at the N-terminal SFNF*PQITL and C-terminal TLNF*PISP natural sites of the protease precursor**

Compared to mature protease PR100, PR107 precursor contained the additional N-terminal heptapeptide GTVSFNFW present in the pol reading frame (Fig. 1a). To determine the specificity of cleavage at the N-terminal end of the protease precursor PR107 and at both the N and C termini of the β-gal-fused protease precursor PR107-β-gal expressed in E. coli, the 11K band produced by the PR107 clone and the 11K and 116K bands from the PR107-β-gal fusion construct were isolated and sequenced. The first six cycles showed the expected sequence P-Q-I-T-L-W for the two 11K proteases and P-I-S-P-G-G for the 116K processed β-gal. This confirmed that protease could mature from its PR107 precursor, by cleavage at the N-terminal site and the protease-β-gal junction, both cleavages occurring at specific F-P bonds.

**Phenotype of mutants in the N- and C-terminal autoprocessing sites**

Since any mutation upstream of the F-P scissile bond at the protease-β-gal junction or downstream of the F-P bond in the N-terminal site would alter the protease sequence and could therefore affect its enzymic activity, we restricted our mutational analysis to residues downstream of the C-terminal automaturation site. The P1' and P2' positions have been extensively studied (Margolin et al., 1990; Partin et al., 1990; Tritch et al., 1991), whereas the requirements for the P3', P4' and P5' subsites have not yet been clearly defined. The pattern of protease release by the RR-β-gal mutant, as shown in Fig. 4(b), was similar to its wild-type PR107-β-gal equivalent (Fig. 3b), although a slightly delayed processing (0.5 to 1 h) could be observed. This suggested that the presence of the two positively charged arginine residues at subsites P4' and P5' in the C-terminal processing site did not significantly impair the F-P scissile bond cleavage.

To analyse the influence of the amino acid residue at position P3' in the N-terminal site on protease autoprocessing, the sequence TLNF*PQITL, consensus to the natural sequence SFNF*PQITL found at the N terminus of the protease, was introduced at the PR107-β-gal junction. We constructed an appropriate vector, termed PRamb3'LZ, which expressed a β-gal-fused protease in which the protease sequence was flanked by two consensus N-terminal processing sites, i.e. its natural upstream SFNF*PQITL site and a downstream site with an amber stop codon in place of the isoleucine codon (TLNF*FQambTL; Fig. 1a). Any substitution at the P3' subsite, downstream of the F-P bond, would therefore not affect the protease sequence. The E. coli strain harbouring PRamb3'LZ was then transformed by each of 12 plasmids expressing an amber suppressor tRNA, and protease autoprocessing at the protease-β-gal junction was assayed by the occurrence of the 116K β-gal band released from the 127K protease-β-gal fusion in SDS-PAGE and immunoblotting.

As shown in Table 2, the efficacy of suppression, determined from the β-gal activity, was found to vary from 0-2% (lysine) to 41% (glycine) for the different suppressor tRNAs, a result which confirmed previous studies (Kleina et al., 1990b). Protease autoprocessing was therefore analysed at late times of arabinose induction (4 h), to compensate for the low level of β-gal expression obtained with certain suppressors. Due to amino acid mischarging by certain suppressor tRNAs (Normanly et al., 1990), some amino acid substitutions
Table 2. Phenotype of amino acid substitutions at position P3' in the TLNF*PQambT autoprocessing site placed at the protease–β-gal junction*

<table>
<thead>
<tr>
<th>Suppressor tRNA</th>
<th>β-gal activity (U/ml)</th>
<th>Suppression efficiency† (%)</th>
<th>Autoprocessing‡ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR107–β-gal alone</td>
<td>3631 (100%)</td>
<td>–</td>
<td>100</td>
</tr>
<tr>
<td>PRamb3'FLZ alone</td>
<td>&lt;2</td>
<td>0-0</td>
<td>–</td>
</tr>
<tr>
<td>PRamb3'FLZ co-expressed with Gly2 (62% G; 37% E)</td>
<td>1489</td>
<td>41-0</td>
<td>70-75</td>
</tr>
<tr>
<td>HisA (94% H)</td>
<td>803</td>
<td>22-1</td>
<td>70-75</td>
</tr>
<tr>
<td>Thr2 (8% T; 86% K)</td>
<td>172</td>
<td>4-7</td>
<td>100</td>
</tr>
<tr>
<td>Ala2 (97% A)</td>
<td>159</td>
<td>4-3</td>
<td>100</td>
</tr>
<tr>
<td>GluA (59% E; 17% Q; 6% Y; 6% R)</td>
<td>129</td>
<td>3-5</td>
<td>100</td>
</tr>
<tr>
<td>Phe (98% F)</td>
<td>91</td>
<td>2-5</td>
<td>100</td>
</tr>
<tr>
<td>Arg (91% K; 5% R)</td>
<td>46</td>
<td>1-2</td>
<td>100</td>
</tr>
<tr>
<td>Pro (85% P)</td>
<td>38</td>
<td>1-0</td>
<td>60-70</td>
</tr>
<tr>
<td>Val (84% K; 5% V)</td>
<td>27</td>
<td>0-7</td>
<td>100</td>
</tr>
<tr>
<td>Cys (90% C)</td>
<td>23</td>
<td>0-6</td>
<td>100</td>
</tr>
<tr>
<td>Ile2 (93% K)</td>
<td>15</td>
<td>0-4</td>
<td>100</td>
</tr>
<tr>
<td>Lys (90% K)</td>
<td>7</td>
<td>0-2</td>
<td>100</td>
</tr>
</tbody>
</table>

* E. coli cells harbouring PRamb3'FLZ were transformed with pct2 plasmid expressing one of the 12 suppressor tRNA genes listed in the table. The specificity and efficiency of the amino acid insertion by each suppressor is indicated in parentheses (Normanly et al., 1990).
† The degree of amber suppression was estimated by β-gal activity, expressed as units per ml (U/ml; Miller, 1972). Controls were PR107–β-gal (100% control for β-gal activity and autoprocessing) and PRamb3'FLZ alone (negative control for β-gal and suppression activity).
‡ Protease autoprocessing was estimated by scanning of the 127K/116K doublet bands of protease-fused and protease-released β-gal in immunoblots of SDS-polyacrylamide gels.

were under-represented (Table 2). Only three amino acids substituting for isoleucine at position P3' were found to have some discrete deleterious effect on the protease autoprocessing efficiency. Glycine, histidine (94% H inserted) and proline (85% P inserted) showed a slight but significant delay in protease processing. Gly2 tRNA inserted glycine and glutamic acid with similar efficiency, but comparison with GluA tRNA, which mainly inserted glutamic acid, allowed us to assign the observed effect on processing to the glycine substitution.

Discussion

In the present study we analysed HIV-1 protease autoprocessing using synthetic genes encoding a protease precursor of 107 amino acids (PR107) and the mature form of the enzyme (PR100), unfused and fused to the N terminus of β-gal (PR107–β-gal and PR100–β-gal), or to its C terminus (β-gal–PR107). It has to be kept in mind that the concept of ' autoprocessing' used for retroviral proteases has been based on the experimental observation that wild-type protease is released from protease-containing polyproteins expressed in E. coli, whereas the mutant protease domain is not (Deboutck et al., 1987, 1990; Kraisslich & Wimmer, 1988; Loeb et al., 1989; Mous et al., 1988; Strickler et al., 1989; and refer to Fig. 3). However, such results do not unambiguously prove that the PR domain embedded in a fusion construct is itself an active protease. It cannot be excluded that a low level of cellular proteolytic activity might randomly cleave the fusion protein and lead to release of a low amount of PR, which in turn acts in trans to generate further PR.

Yields of protease and β-gal were found to be significantly higher when the gene for PR107 was fused to the 5' end of the E. coli lacZ gene (PR107–β-gal), and comparison of the level of expression of unfused and β-gal-fused mature forms (PR100 and PR100–β-gal) suggested that the enhancing effect was post-transcriptional and due to the additional 5' sequence of the PR107 gene. In contrast to a previous report which showed an inactive form of β-gal-fused protease accumulating in inclusion bodies in E. coli (Giam & Boros, 1988), all our cloned proteases (except G33 and G33–β-gal protease mutants) were recovered in an active form under mild, physiological conditions, and none of them, not even the inclusion-forming PR107–β-gal, required a solubilization step in urea-containing denaturing buffer. Our data are therefore reminiscent of previous observation on high protease yields obtained with certain clones of protease precursors encoding N-terminal sequences of 11 or 56 residues upstream of Pro (1) (Deboutck et al., 1990). According to our results, an upstream sequence of seven
residues is sufficient to enhance the protease production without altering protease activity.

Upstream fusion to β-gal, as in PR107-β-gal, did not result in any detectable cytotoxicity for E. coli cells (Fig. 6), and a comparison of PR107- and PR107-β-gal-expressing clones with the inactive mutants G33 and G33-β-gal suggested that the PR107- or G33-induced c.p.e. resulted from the expression of a foreign gene product in E. coli rather than to the enzymatic properties of the protease itself. Only the two upstream fusion constructs PR107-β-gal and G33-β-gal gave rise to intracellular inclusions. This is in contrast to previous reports in which intracellular precipitates were only observed with mature protease or with protease precursors containing both upstream and downstream sequences expressed in heat-shock response-deficient E. coli (Debouck et al., 1990). Immunoelectron microscopy of the PR107-β-gal inclusions revealed that protease remained associated with the inclusion bodies, even after processing (Fig. 7), a phenomenon which could account for the absence of toxicity observed with our β-gal-fused protease gene constructs.

Processing at Phe-Pro bonds occurred in vivo with a similar efficiency at the PR107-β-gal and PR100-β-gal junctions, at the N terminus of the PR107 precursor and at the N and C termini of the β-gal-fused protease precursor PR107-β-gal. The N-terminal site was also correctly processed when placed at the protease C-terminal end. This suggests that each autoprocessing site contains its own information for cleavage, independent of protein context, and supports the hypothesis that protease autoprocessing occurs as a result of intermolecular interactions (Miller et al., 1990). However, this does not imply that the N- and C-terminal processing are independent events, since the assay used cannot distinguish independence of these two reactions from a situation in which the C-terminal cleavage is dependent on the N-terminal cleavage, but the latter one occurs at a higher rate (Kraüsslich et al., 1989; Strickler et al., 1989). Moreover, a recent study using mutated fusion proteins expressed in E. coli indicated that altering one of the protease cleavage sites influences the cleavage at the non-mutated site (Louis et al., 1991).

Upon substitution of the P-G dipeptide at positions P4' and P5' in the C-terminal site by the two positively charged amino acids R–R, protease release was only slightly delayed. This result was not surprising if one considers (i) the topography of residues P4' and P5', positioned downstream of the F–P scissile bond and outside of the active site cleft, (ii) the tolerance of an arginine residue at the P4' position of the natural site at the p9NC®–p6LI®° junction (Debouck et al., 1990) and (iii) the tolerance of an arginine residue at P1' position in non-viral protein substrates of HIV-1 protease, e.g. troponin C and calmodulin (reviewed in Tomasselli et al., 1991).

The upstream and downstream boundaries of HIV-1 protease constitute its two cleavage sites, SFNF*PQIT at its N terminus and TLNF*PISP at its C terminus, respectively. In the present study, we have focused on the subsite P3' of the N-terminal autoprocessing site of the protease for the two following reasons. (i) It has been previously shown that cleavage at the F–P bond constituting the N terminus of the mature protease takes place significantly faster than cleavage at its C-terminal F–P bond (Strickler et al., 1989), and that the natural N-terminal autoprocessing site VSFNF*PQITL had the highest Vmax/Km ratio (Kraüsslich et al., 1989). (ii) As a result of mutational analyses of native substrates of HIV-1 protease (Partin et al., 1990; Tomasselli et al., 1991; Tritch et al., 1991) and of its autoprocessing sites (LeGrice et al., 1988; Loeb et al., 1989; Louis et al., 1991), and from theoretical considerations on the structure of the protease (Hellen et al., 1989; Swanstrom et al., 1989), the following consensus sequence for an autoprocessing site has been proposed: P4 (small and hydrophobic), P3 (undefined), P2 (small), P1 (aromatic or large and hydrophobic), P1' (proline), P2' (small and hydrophobic), P3' (variable).

Amino acid residues at subsites P3 and P3' have been postulated to be critical for the precise alignment of a peptide substrate within the protease active site cleft (Sali et al., 1989; Miller et al., 1990). In addition, once such a ligand is positioned within the cleft, P3 and P3' are adjacent to both flaps of the protease dimer (Erickson et al., 1990; Harte et al., 1990; Lapatto et al., 1989; Miller et al., 1989a, b; Moore et al., 1989; Navia et al., 1989; Suguna et al., 1987; Weber et al., 1989; Wlodawer et al., 1989). However, no experimental data are available on the influence of the residue at the P3' subsite of the N-terminal site on protease autoprocessing, due to the fact that substitutions at this P3' position would change the protease N-terminal sequence and therefore possibly alter its proteolytic activity. This was the case when aspartic acid at P3' in the N-terminal site of the protease was substituted for isoleucine: no self-processing was observed with the mutant, whereas cleavage could be rescued in trans by the wild-type protease (Partin et al., 1990).

For an indirect analysis of the effect of P3' substitutions in the upstream site on protease autoprocessing, we introduced a consensus N-terminal processing site at the C-terminal extremity of the protease (at its junction with β-gal), and substituted a suppressible amber mutant codon for the isoleucine codon at position P3' (TLNF*PQambT). When the amber mutation was assayed for rescue by a series of 12 suppressor tRNAs, only G, H and P substituting for I were found to reduce
the processing efficiency (Table 2). This suggested a tolerance for a wide variety of amino acid residues at P3' of the N-terminal processing site of HIV-1 protease, except for amino acid substitutions with strong effects on the polypeptide chain secondary structure, e.g. glycine or proline. This implied a relatively high degree of flexibility of the protease flaps in their contact with the amino acid residue at P3'. The design of more efficient HIV protease inhibitors should take into account all available data on protease cleavage of viral and non-viral natural substrates (Rivière et al., 1991), but the primary and therefore essential event is indeed its own cleavage and release by an autoprocessing mechanism.

The authors are grateful to Dr Edmund Cheng for providing them with anti-protease antibodies. This work has been supported by the Agence Nationale de Recherche sur le SIDA (ANRS, AC-2 and AC-14), the Conseil de la Région Midi-Pyrénées (UMR 5614, Conseil de la Région Languedoc-Roussillon (PB) and the Fondation pour la Recherche Médicale (PB). The original project was launched thanks to grant CRE 88-3001 from the Institut National de la Santé et de la Recherche Médicale.

References

Autoprocessing of HIV-1 protease


(Received 30 July 1991; Accepted 28 October 1991)