Rabbit endogenous retrovirus-H encodes a functional protease

Cécile Voisset,1 Richard E. Myers,1 Alex Carne,2† Paul Kellam1 and David J. Griffiths1

1Wohl Virion Centre, Windeyer Institute of Medical Sciences, University College London, 46 Cleveland Street, London W1T 4JF, UK
2Institute of Cancer Research, Chester Beatty Laboratories, London, UK

Recent studies have revealed that ‘human retrovirus-5’ sequences found in human samples belong to a rabbit endogenous retrovirus family named RERV-H. A part of the gag–pro region of the RERV-H genome was amplified by PCR from DNA in human samples and several forms of RERV-H protease were expressed in bacteria. The RERV-H protease was able to cleave itself from a precursor protein and was also able to cleave the RERV-H Gag polyprotein precursor in vitro whereas a form of the protease with a mutation engineered into the active site was inactive. Potential N- and C-terminal autocleavage sites were characterized. The RERV-H protease was sensitive to pepstatin A, showing it to be an aspartic protease. Moreover, it was strongly inhibited by PYVPheStaAMT, a pseudopeptide inhibitor specific for Mason–Pfizer monkey virus and avian myeloblastosis-associated virus. A structural model of the RERV-H protease was constructed that, together with the activity data, confirms that this is a retroviral aspartic protease.

INTRODUCTION

Rabbit endogenous retrovirus-H (RERV-H) is a recently characterized endogenous retrovirus family of the European rabbit (Oryctolagus cuniculus) (Griffiths et al., 2002). RERV-H was first identified in human tissues by PCR and RT-PCR, and was originally thought to be a novel human retrovirus, denoted human retrovirus-5 (HRV-5) (Griffiths et al., 1997). However, despite having many characteristics of infectious retroviruses, in particular an apparent association with inflammatory diseases and lymphoma (Rigby et al., 1998; Brand et al., 1999; Griffiths et al., 1999; Murovska et al., 2000; Kozireva et al., 2001), cloning of the HRV-5 genome and of some viral integration sites revealed the true origin of this virus and led to its re-designation as RERV-H (Griffiths et al., 2002).

The RERV-H family appears to be restricted to the European rabbit because the genomes of related members of the order Lagomorpha, including hares and pikas, do not contain RERV-H (Griffiths et al., 2002). This observation, together with the conservation of open reading frames (ORFs) for gag, pro and pol, indicates that this virus entered the rabbit genome relatively recently in evolution and raises the possibility that it is still active and infectious. Since its first discovery in human samples, we have been studying the function of several RERV-H proteins in vitro and here we describe our studies on the RERV-H viral protease (PR).

Retroviral PRs belong to the family of aspartic proteases, which has many members both in eukaryotes and prokaryotes (Davies, 1990; Hill & Phylip, 1997). This family is characterized by sequence conservation around the active site, an acid pH optimum and similarity of three-dimensional structure (Wlodawer & Gustchina, 2000). The cellular aspartic proteases are bilobal in structure, with two domains of similar amino acid sequence in a single polypeptide (Davies, 1990). In contrast, retroviral PRs are dimeric enzymes composed of two separate but identical subunits (Bianchi et al., 1990; Katoh et al., 1985). The N- and C-terminal amino acids of each subunit are essential for dimer stability (Miller et al., 1989; Wlodawer et al., 1989).

Among retroviruses, several strategies are used to control expression of PR (Hatfield et al., 1992). In common with other betaretroviruses, the PR of RERV-H is encoded within a separate ORF (the pro gene), in fusion with an N-terminal deoxyuridine triphosphatase (dUTPase, DU) domain (Elder et al., 1992) and its expression requires a ribosomal frameshifting event (Hatfield et al., 1992). Thus, the PR domain is contained in both the Gag–Pro and Gag–Pro–Pol precursors. Mature retroviral PRs are able to autocatalyse their own release from these polyproteins either during or shortly after assembly and budding. They also cleave the various Gag polyprotein precursors to yield the mature viral proteins as a late event in assembly. These tightly regulated proteolytic events result in morphological changes in the
virus particle that are essential for virus infectivity (Kohl et al., 1988; Sommerfelt et al., 1992). Here, we show that RERV-H PR cloned from human DNA and expressed in a prokaryotic expression system is active on RERV-H proteins and is sensitive to known inhibitors of retroviral PRs.

**METHODS**

**PCR amplification of RERV-H NC-DU-PR region.** Genomic DNA was prepared as previously described (Griffiths et al., 2002).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCF1</td>
<td>ATCTGTAGGGATGTCATGGATGATG</td>
<td>100</td>
</tr>
<tr>
<td>PRR1</td>
<td>ATCACGAATTATTGGCGTATTCCGC</td>
<td>200</td>
</tr>
<tr>
<td>NCF2</td>
<td>CTGGATCACATGCTGCAGTAGCTG</td>
<td>400</td>
</tr>
<tr>
<td>PRR2</td>
<td>TGATGGTTGCAAATGGCCTGCCTC</td>
<td>39</td>
</tr>
<tr>
<td>NCF2</td>
<td>CTGGATCACATGCTGCAGTAGCTG</td>
<td>72</td>
</tr>
<tr>
<td>PRR2</td>
<td>TGATGGTTGCAAATGGCCTGCCTC</td>
<td>139</td>
</tr>
<tr>
<td>NCF2</td>
<td>CTGGATCACATGCTGCAGTAGCTG</td>
<td>121</td>
</tr>
<tr>
<td>PRR2</td>
<td>TGATGGTTGCAAATGGCCTGCCTC</td>
<td>172</td>
</tr>
<tr>
<td>NCF2</td>
<td>CTGGATCACATGCTGCAGTAGCTG</td>
<td>106</td>
</tr>
<tr>
<td>PRR2</td>
<td>TGATGGTTGCAAATGGCCTGCCTC</td>
<td>105</td>
</tr>
</tbody>
</table>

The NC-DU-PR region of RERV-H was amplified by nested-PCR. First stage reaction mixtures contained 10 pmol of primers NCF1 (ATCTGTAGGGATGTCATGGATGATG) and PRR1 (ATCACGAATTATTGGCGTATTCCGC), 250 μM of each dNTP (Pharmacia), 200 ng of DNA and 1-25 U of Pfu Turbo DNA polymerase (Stratagene) with the supplied reaction buffer in a final volume of 50 μl. Cycling conditions were 40 cycles of 94 °C, 70 s; 58 °C, 70 s; 72 °C, 2 min with an initial denaturation at 94 °C for 4 min. One μl of each first-stage product was transferred to a second PCR reaction containing NCF2 (CTGGATCACATGCTGCAGTAGCTG) and PRR2 (TGATGGTTGCAAATGGCCTGCCTC) and amplified further.
Molecular cloning. NC-DU-PR PCR products were first blunt-end cloned into EcoRV-digested pBlueScript KS(–) and then subcloned into BamHI/HindIII-digested pAlter-1 (Promega) using sites engineered into the PCR primers, to generate plasmid pNC-PR (Fig. 1a). The PR region (nt 717–1283) was obtained by PCR and subcloned into the bacterial expression vector PET21b (Novagen) in-frame with the C-terminal hexahistidine tag to create plasmid pPR(wt). pNC-PR was mutated (nt 799–800, Fig. 1a) to substitute the active site aspartine with alanine using the oligonucleotide 5'-ATTATACTTACATCTGCACCTGTGGCTAGCAGGCCCTG-3' and the Altered sites II mutagenesis system (Promega, as recommended), to generate plasmid pNC-PR(D172A). This mutated PR fragment was subcloned into PET21b as above to create pPR(D172A).

Plasmids pShort-PR(wt) and pShort-PR(D172A) were derived from pNC-PR(wt) and pNC-PR(D172A) respectively by amplifying the fragments corresponding to nt 717–1025 in Fig. 1(a) and cloning into EcoRI-digested PET21b using sites engineered into the PCR primers.

Plasmids pPR160, pPR160N and pPR160C were created by subcloning fragments of the original HRV-5/RERV-H PR clone (Griffiths et al., 1997; GenBank acc. no. U46939; corresponding to amino acids 173–332, 173–255 and 247–332 of Fig. 1a respectively) into EcoRI-digested PET21b in-frame with a C-terminal c-Myc tag (EQKLISEEDL) engineered into the reverse PCR primer. The RERV-H gag gene obtained from plasmid pHRV5.6 (Griffiths et al., 2002) was cloned into pGeX6p1 (Pharmacia) in fusion with the N-terminal glutathione S-transferase (GST) domain and with C-terminal c-Myc and hexahistidine tags introduced by PCR to create plasmid pGex-Gag.

Protein expression and analysis. Escherichia coli strain BL21-CodonPlus(DE3)-RIL (Stratagene) was used as the host for all expression experiments. Transformed bacteria containing each expression plasmid were grown overnight at 37 °C in Luria–Bertani broth (LB) containing 100 μg ampicillin ml⁻¹ and 30 μg chloramphenicol ml⁻¹. Bacteria were then diluted 1:10 in fresh LB with antibiotics and grown for 1 h at 37 °C before protein expression was induced with 1 mM IPTG for 3 h at 37 °C.

Proteins were separated by SDS-PAGE and transferred to PVDF membrane (Amersham) using established procedures (Sambrook et al., 1989). Monoclonal antibodies (mAbs) to RERV-H PR (mAb-Pro1 and mAb-Pro9) were generated in rats as described (Dean et al., 1986; Griffiths, 1996). Polyclonal anti-RERV-H Gag antibodies raised in rabbits were a kind gift from John Hackett, Abbott Laboratories. RERV-H PR was detected by rat anti-PR mAbs or the mouse anti-c-Myc mAb 9E10 (Evan et al., 1985; obtained from the Institute of Cancer Research, Sutton, UK). The secondary antibodies were goat anti-rat IgG, goat anti-mouse IgG or goat anti-rabbit Ig conjugated to horseradish peroxidase (Harlan). Bound antibodies were detected by enhanced chemiluminescence using the ECL Western blotting detection reagents (Amersham) as recommended.

Affinity purification of recombinant proteins. Cultures (500 ml) of E. coli BL21-CodonPlus(DE3)-RIL transformed with either pGex-Gag or pPR(wt) were grown at 37 °C in LB with antibiotics to an optical density at 600 nm of 0.6. IPTG was added to 1 mM and cells were grown for a further 3 h before harvesting by centrifugation (4000 g, 10 min, 4 °C). The pellets were suspended in 5 ml of buffer T (100 mM NaH2PO4 pH 8.0, 10 mM Tris/HCl pH 8.0, 6 M guanidium. HCl) per gram of bacterial pellet (2-6 g for 500 ml culture) and shaken gently for 1 h at room temperature. The bacterial lysate was cleared by centrifugation (10000 g, 30 min, room temperature) and the supernatant was added to 3 ml of 50 % Ni–NTA resin (Qiagen, pre-equilibrated in buffer L) and mixed gently for 1 h at room temperature. The resin was pelleted by centrifugation, resuspended in 10 ml of buffer D (100 mM NaH2PO4 pH 8-0, 10 mM Tris/HCl pH 8-0, 8 M urea, 20 mM imidazole) and mixed for 1 h at room temperature before loading into a column and washing with 10 ml of buffer D. Bound proteins were eluted with four 1 ml aliquots of buffer D containing either 250 mM imidazole [PR(wt)] or 175 mM imidazole (GST–Gag). Purified PR(wt) proteins (800 μg in 500 μl of buffer D) were dialysed against 4 l of distilled water for 15 h at 4 °C and then for 6 h against 100 ml of activity buffer (5 mM DTT, 250 mM NaCl, 2 mM PMSF, 2 mM EDTA, 100 mM sodium phosphate pH 5-3) at 4 °C in a mini slide with a 3500 Da cut-off (Pierce).

Protease activity. For analysis of autoprocessing, lysates were prepared from 3 ml of IPTG-induced bacterial culture. Cells were pelleted by centrifugation (4000 g, 4 °C, 10 min) and suspended in 400 μl of buffer T (20 mM Tris/HCl pH 8-0, 6 mM MgCl2, 1 mM EDTA, 1 % Triton X-100, 10 % glycerol). The resulting lysates were incubated for 10 min on ice with 0.1 units Benzonase (Merck) ml⁻¹, loading buffer was added (5× is 250 mM Tris/HCl pH 8-0, 5 % β-mercaptoethanol, 20 % glycerol, 5 % SDS, 0.5 % bromophenol blue) and the extracts analysed by SDS-PAGE and immunoblotting.

For assessing PR cleavage of RERV-H Gag polyproteins, 1-5 ml of E. coli culture expressing the PR clone under test (cultured and induced as above) was mixed with 0.5 ml of bacteria expressing GST–Gag. This mixture was pelleted, suspended in 250 μl of activity buffer and sonicated three times for 30 s on ice. Bacterial lysates were incubated for 10 min on ice with Benzonase (Merck) and incubated for 3 h at 37 °C before analysing by SDS-PAGE and immunoblotting.

Protease assays on GST–Gag proteins eluted from SDS-PAGE gels were performed with GST–Gag that had been affinity purified on Ni–NTA and separated on a 10 % polyacrylamide gel. After SYPRO-Orange staining the gel was imaged and the protease activity quantified as described (Appendix 2).
staining (Bio-Rad), a piece of gel containing the full-length 95 kDa GST–Gag protein was excised, finely sliced and transferred to a tube containing 300 μl of activity buffer. The gel fragments were centrifuged (14 000 g, 4 °C, 10 min) and the supernatant containing the GST–Gag eluted protein was retained. As a control, a gel piece containing no protein was treated in the same way. Bacterial cultures (3 ml) expressing either PR(wt) or PR(D172A) were centrifuged and the pellet suspended in 400 μl of activity buffer before sonication and treating the resulting lysates with Benzonase. GST–Gag proteins eluted from an SDS-PAGE gel (100 μl) were mixed with 100 μl of PR(wt) or PR(D172A) bacterial lysate and incubated for 3 h at 37 °C.

The activity of purified and dialysed RERV-H PR was assayed on GST–Gag. Cells pelleted from 3 ml of culture expressing GST–Gag were suspended in 400 μl of activity buffer and sonicated. The lysates were treated with Benzonase and a 100 μl aliquot was mixed with 100 μl of purified and dialysed RERV-H PR(wt) or PR(D172A), and incubated for 3 h at 37 °C.

**N-terminal microsequencing.** Purified and dialysed PR(wt) proteins were separated by SDS-PAGE and transferred to PVDF membrane (Immobilon-P, Millipore) in transfer buffer (10 mM Tris/ HCl, pH 8-3, 100 mM glycine, 15 % methanol) for 90 min at 230 mA. Following transfer, filters, were stained (0-1 % Coomassie Brilliant blue R250 in 40 % methanol), destained in 50 % methanol and rinsed several times in water. Bands were excised and sequenced using an Applied Biosystems 477 pulse-liquid protein sequencer with on-line phenylthiohydantoin analysis.

**Inhibitor assays.** Inhibition assays were performed as protease assays, with addition of increasing amounts of pepstatin A (Roche) or PYVPheStaAMT (kindly provided by Jan Konvalinka) solubilized in DMSO. Whatever the inhibitor concentration, each assay contained the same concentration of DMSO (9-4 %). Products were detected after immunoblotting with an anti-c-Myc mAb.

**Sequence analysis.** DNA sequencing was performed using a Beckman–Coulter CEQ2000 automated DNA sequencer. Computer-aided analysis of protein and nucleotide sequences was performed with the Sequencher program (Gene Code). The program CORA (Orengo, 1999) was used to construct the structural alignment of retroviral PRs and this alignment was then used to generate the theoretical structural model of the RERV-H PR with Modeller 4.0 (Sali & Blundell, 1993) using the human immunodeficiency virus type-1 (HIV-1) and Rous sarcoma virus (RSV) PR monomers as templates.

**RESULTS**

**Amplification of RERV-H pro from human DNA samples**

A nested-PCR primer set was designed to amplify the NC-DU-PR region of RERV-H. Amplification conditions were optimized to give a sensitivity of detection in titration experiments of about 10 copies of RERV-H plasmid. Of 30 human DNA samples tested, the desired 1·4 kb fragment was amplified from the genomic DNA of only one blood donor (sample A) and from one patient with ulcerative colitis (sample B) (data not shown). Both samples were previously shown to have a relatively high load of RERV-H sequence (Griffiths et al., 2002) and were positive after amplification with the first stage PCR primers. Sequencing of two plasmid clones from each patient revealed strong similarity (96 % amino acid similarity) with the reference pHHRV-5.6 plasmid sequence (GenBank acc. no. AF480924; Griffiths et al., 2002). In subsequent experiments, we analysed the PR activity of one clone from each patient. These two proteins had identical behaviour and so only the activity of the PR amplified from sample A is described here (Fig. 1a).

**Autoprocessing activity of the RERV-H protease**

The activity of the RERV-H PR was studied in *E. coli* using a variety of expression constructs (Fig. 1b). To determine whether the recombinant RERV-H PR is active, we first tested its ability to cleave itself out of a larger precursor molecule since this is a required function for retroviral PRs (Wan et al., 1996). Because we were not certain of the termini of the mature RERV-H PR, we expressed a fragment, denoted PR(wt), that begins 27 residues upstream of the DTG active site (estimated from a structural analysis of retroviral PRs, see below) and ends at the natural stop codon of the pro gene (Fig. 1a, b). To facilitate affinity purification on nickel–agarose beads, the PR(wt) protein was expressed in fusion with a C-terminal hexahistidine tag. The tagged PR(wt) is a 222 amino acid protein with a calculated molecular mass of 24 665 Da.

Expression of the PR(wt) polypeptide was induced in *E. coli* and examined by immunoblotting using two mAbs raised against a fragment of RERV-H PR. These are mAb-Pro9, which recognizes a central epitope on RERV-H PR, and mAb-Pro1, which binds a C-terminal epitope (Fig. 1c). As shown in Fig. 2, mAb-Pro9 detected five prominent bands at approximately 24, 22, 20, 18 and 16 kDa while bands of 24, 22 and 12 kDa and doublets of 17 and 15 kDa were detected with mAb-Pro1.

![Fig. 2. RERV-H PR autoprocessing. RERV-H PR(wt) and PR(D172A) were expressed in E. coli and lysates analysed by immunoblotting and probing with mAb-Pro9 (left panel) and mAb-Pro1 (right). Arrowheads show the molecular masses of the detected proteins. Specific processing of the precursor is observed with PR(wt) but not PR(D172A).](image-url)
In order to discriminate authentic RERV-H PR autoprocessing from proteolysis due to bacterial proteases, we mutated the active site of PR(wt) to create PR(D172A). This mutation is predicted to inactivate the RERV-H PR without affecting the susceptibility of the precursor protein to bacterial proteolysis (Darke et al., 1989; Kohl et al., 1988). Where the mutant PR was expressed, the 25 kDa full-length protein was observed using both mAbs, with additional bands at 20 and 12 kDa also detected by mAb-Pro1 (i.e. C-terminal fragments) (Fig. 2). These are most likely due to cleavage by bacterial enzymes or to non-specific cleavage during sample handling. MAb-Pro9 detected the 25 kDa full-length polypeptide only. Taken together, these results indicate that RERV-H PR is able to autocleave itself from a larger polyprotein and that the 24, 22, 18 and 16 kDa processed proteins detected by mAb-Pro9 on PR(wt) most likely result from the autocatalytic processing of the RERV-H PR itself.

**Processing of RERV-H Gag polyprotein**

We next wished to establish whether RERV-H PR expressed in bacteria is capable of processing its cognate Gag protein in trans. We successfully expressed a 95 kDa recombinant RERV-H Gag polyprotein fused with GST at the N terminus and C-terminal hexahistidine and c-Myc tags (GST–Gag; Fig. 1b). Note that the GST moiety was introduced to increase the level of expression of the full-length Gag and was not used to purify the protein on glutathione–agarose.

When GST–Gag was expressed in *E. coli*, the largest band detected on immunoblots with the anti-c-Myc mAb was 95 kDa, which is the expected size of the full-length GST–Gag precursor (Fig. 3a). Bands at approximately 65, 45, 42, 30 and 20 kDa were also observed and these may represent partial degradation products of the Gag precursor, or may be the result of internal translation initiation. The instability of retroviral Gag polyproteins expressed in *E. coli* has been described previously (Debouck et al., 1987; Graves et al., 1988; Hansen et al., 1988; Mueller-Lantzsch et al., 1993).

As shown in Fig. 3(a), incubation of GST–Gag with a PR(wt) bacterial lysate resulted in the appearance of major cleavage products of 28, 18, 15 and 13 kDa, the disappearance of the 95, 65 and 45 kDa proteins, and an increase in the amount of the 30 and 42 kDa species. In contrast, no additional cleavage of Gag was observed when GST–Gag was exposed to a PR(D172A) bacterial extract. Using an

---

**Fig. 3.** RERV-H GST–Gag precursor processing in trans by RERV-H PR. (a) Processing of GST–Gag bacterial lysate. Bacterial lysates containing GST–Gag precursor, PR(wt) and PR(D172A) polypeptides were mixed and incubated and products analysed by immunoblotting and probed with anti-c-Myc and anti-CA antibodies. (b) Processing of the GST–Gag precursor previously eluted from a SDS-polyacrylamide gel. The eluted GST–Gag precursor was incubated with PR(wt) and PR(D172A) bacterial lysates and cleavage products analysed by immunoblotting using the anti-c-Myc mAb. (c) Purified PR(wt) proteins were dialysed and PR activity was assayed on the GST–Gag precursor and products analysed by immunoblotting using the anti-c-Myc mAb. Arrowheads show the molecular masses of the detected proteins.
Fig. 4. Characterization of C-terminal autoprocessing site. Proteins were affinity purified from bacterial lysates expressing PR(wt) on nickel–agarose under denaturing conditions. (a) Coomassie-stained gel of purified and dialysed PR(wt) proteins. N-terminal sequences of the boxed bands are shown. (b) Activity of the 103 residue shortPR (residues 145–247, Fig. 1a), assessed on the GST–Gag polyprotein by immunoblotting with the anti-c-Myc mAb. Arrowheads show the molecular masses of the detected proteins.

Fig. 5. Inhibition of RERV-H PR. GST–Gag precursor, PR(wt) and PR(D172A) polypeptides were expressed in E. coli and bacterial lysates were mixed in the presence of increasing amounts of (a) pepstatin A or (b) the pseudopeptide inhibitor PYVPheStaAMT. Products were detected by immunoblotting and staining with the anti-c-Myc mAb.
Characterization of the RERV-H PR autoprocessing sites

In order to identify RERV-H PR autoprocessing sites and define the limits of the mature PR, N-terminal amino acid sequencing was performed on the 24, 17 and 15 kDa purified His-tagged PR(wt) polypeptides. The N terminus of the 24 kDa polypeptide was found to be AALS (band 1, Fig. 4a). Thus, the 24 kDa protein observed is not the complete polypeptide encoded by plasmid pPR(wt) but an N-terminal autoprocessed intermediate. This most likely explains the size difference observed on SDS-PAGE between PR(wt) and PR(D172A) since PR(D172A) is not autoprocessed.

We determined the N-terminal sequences of the 17 and 15 kDa fragments on the basis that this would provide information on the C-terminal autoprocessing site of the mature RERV-H PR (Fig. 4a) The N terminus of the 17 kDa band (band 2) and the upper band of the 15 kDa doublet (band 3) are located upstream of the GRD motif (position 239–241, Fig. 1a). As this motif is essential for retroviral PR activity, the PR fragments released by cleavage at these sites are probably inactive. Indeed, whether cleavage at these sites is actually catalysed by RERV-H PR is unclear. It is possible that these cleavage sites become inaccessible to bacterial proteases following conformational changes that occur after the authentic N-terminal autoprocessing step. The N terminus of the lower band of the 15 kDa doublet (band 4, Fig. 4a) is located four amino acids downstream of the GRD motif. Although a PR polypeptide with this C terminus would retain the GRD motif, it would still be seven residues shorter than lentiviral PRs and eight residues shorter than the RSV PR (see below). Since the C-terminal region is important for dimer stability, it appears unlikely that this is the genuine C terminus of RERV-H PR. However, this is the only cleavage site we determined that is downstream of the GRD motif and therefore it is our best candidate for the genuine C terminus of the mature RERV-H PR. The N terminus of the 12 kDa band could not be determined. It should also be noted that we cannot be certain that a 100 amino acid PR (predicted by an N-terminal AALS and C-terminal ILEQ) actually exists because these two cleavage sites were sequenced from two different autoprocessing intermediates.

To determine whether the PR predicted from microsequencing experiments was active, we expressed a 103 amino acid form of the protein (Short-PR, amino acids 145–247, see Fig. 1a, b) in E. coli and analysed its processing activity on the GST–Gag substrate. As shown in Fig. 4(b), the shortPR(wt) is active, but not as much as PR(wt) since the only cleavage product obtained was 42 kDa.

The reduced activity of the 103 residue shortPR(wt) could be due either to a truncated N or C terminus. For human T-lymphotropic virus type I and HIV-1, the C-terminal amino acids are required for enzymatic activity because their involvement in the beta-sheet stabilizes the structure of the
dimer (Hayakawa et al., 1992; Zhang et al., 1991). If this were the case for RERV-H PR, the truncated C terminus (compared to HIV) would be expected to result in reduced dimer stability and less activity. Similarly, it has been shown that additional or missing amino acids at the N terminus of PR are deleterious to the activity of PRs of avian retroviruses (Grinde et al., 1992; Pichova et al., 1992; Sellos-Moura & Vogt, 1996).

**Inhibition of RERV-H PR activity**

The effect of two aspartic protease inhibitors on RERV-H PR activity was assessed. We first used pepstatin A, a broad spectrum inhibitor of both retroviral and cellular aspartic proteases (Nam & Hatanaka, 1986; Ratner et al., 1985). Processing of the GST–Gag precursor by RERV-H PR (wt) was partially inhibited by 0-1 mM pepstatin A and was completely inhibited at 1-5 mM (Fig. 5a). The concentration of pepstatin A required to give inhibition of activity was several orders of magnitude higher than the concentration required for inhibition of cellular aspartic proteases such as pepsin and renin (Umezawa et al., 1970). However, this requirement for a high pepstatin A concentration agrees with results previously reported for retroviral PRs (Dreyer et al., 1989; Katoh et al., 1987).

We next used a pseudopeptide PR inhibitor (PYVPheStaAMT) known to be active against Mason–Pfizer monkey virus (MPMV) and avian myeloblastosis-associated virus (Hruskova-Heidingsfeldova et al., 1995). As with pepstatin A, cleavage of the GST–Gag precursor by RERV-H PR (wt) was inhibited but in this case complete inhibition occurred with 3 nM of PYVPheStaAMT (Fig. 5b). This is similar to the sensitivity of MPMV PR to this inhibitor.

**Modelling of RERV-H PR structure**

The amino acid sequence of the RERV-H PR itself provides strong evidence that it encodes an aspartic protease, since it contains the canonical active site sequence motif (LXDTGAD) between positions 23 and 29 (Fig. 6a, HIV-1 numbering) and a second conserved motif, GRD/N, at position 86–88 (Fig. 6a). Moreover, the spacing between these two motifs is consistent with other retroviral PRs. To complement the biochemical analysis, we sought to derive a structural model for the RERV-H PR from the structures of those retroviral PRs that have been crystallized.

The RERV-H PR amino acid sequence was compared to five retroviral PRs with structures in the Protein Database (Berman et al., 2000), namely HIV-1 (1hx; Krohn et al., 1991), simian immunodeficiency virus (SIV; 1tcw; Hoog et al., 1996), feline immunodeficiency virus (FIV; 4iv; Kervinen et al., 1998), equine infectious anaemia virus (EIAV; 1fmb; Gustchina et al., 1996) and RSV (1bai; Wu et al., 1998). The sequence identity between RERV-H and other retroviral PRs is low (~30% with HIV-1 and RSV) but the predicted structural features are conserved (Fig. 6a).

In particular, residues forming part of the substrate-binding pocket (residues 24–32 and 81–84; HIV-1 numbering) in the other PRs are well conserved in the RERV-H sequence. The flap region corresponding to residues 44–56 is also conserved with flexible glycine residues at positions 49 and 51 (HIV-1 numbering) across all six sequences. The residue at position 50 interacts with the water molecule that is required for substrate cleavage. The RERV-H sequence contains leucine here, which is unusual for this position, although it is also present in the recently characterized murine intracisternal A-particle PR (Strišovský et al., 2002).

Retroviral PRs range in length from 99 (HIV-1) to 125 (RSV) residues, with structural conservation of the N- and C-terminal regions that form the subunit beta-sheet (Rao et al., 1991). On the basis of the structural alignment analysis (Fig. 6a) a three-dimensional atomic model of a 107 residue RERV-H PR was constructed using the program Modeller 4 (Fig. 6b), and this model of RERV-H PR illustrates how the sequence may fold. Regions of sequence conservation from the structural alignment can be spatially arranged such that they form an active site that is consistent with all other retroviral PRs. Existing data on retroviral PRs shows that they exhibit little structural variation around the substrate-binding pocket but have most variation within three external loop structures and this is reflected in the model of RERV-H PR. This model of the RERV-H PR shows that two of the external loops are smaller than those of RSV PR and are similar in size to those of HIV-1 PR. The alignment and modelling of RERV-H PR both show that structural and sequence motifs are conserved with other retroviral PRs and are consistent with the biochemical analysis.

**DISCUSSION**

RERV-H is a recently identified retro-element of rabbits that was originally thought to be an exogenous human retrovirus, denoted HRV-5 (Griffiths et al., 2002). Currently, it is unknown how RERV-H sequences came to be present in human DNA preparations. Laboratory contamination would be one obvious mechanism but cannot explain several observations, notably its disproportionate detection in autoimmune diseases and cancer (Rigby et al., 1998; Brand et al., 1999; Griffiths et al., 1999; Murovska et al., 2000; Kozireva et al., 2001), and its detection by RT-PCR in tissue homogenates separated by sucrose gradient centrifugation (Griffiths et al., 1997).

Zoonotic transfer of RERV-H from rabbits to humans is regarded as unlikely but has not been formally excluded and currently no data exist on the expression or activity of RERV-H proteins. Although some proviral loci are defective, full-length ORFs for gag, pro and pol have been cloned from rabbit and human samples so the possibility remains that some intact loci exist that could encode replication-competent viruses. Since such proviruses have not yet been identified and in the absence of a culture system...
for this virus, we have sought to demonstrate functional activity for the viral enzymes. In this report we have characterized the RERV-H PR. Although the clones used were amplified by PCR from human DNA samples, they do not differ significantly from sequences previously reported in rabbits (Griffiths et al., 2002).

Fig. 6. Structural modelling of RERV-H PR. (a) Structural alignment of RERV-H PR with PRs of HIV-1, SIV, RSV, FIV and EIAV. Residues corresponding to the substrate-binding pocket (p1–p2) and the flap region are indicated. Sequence numbers refer to HIV-1 PR. (b) Theoretical model of RERV-H PR monomer (blue) superimposed with the crystal structure of the RSV PR monomer (grey). The highlighted section (red) shows the active-site triad, DTG. The flap and loop regions are indicated.
Using a prokaryotic expression system to prepare recombinant RERV-H PR enzyme and Gag substrate, we have shown that the RERV-H PR can catalyse the cleavage of both PR-containing substrates (Fig. 2) and RERV-H Gag precursors (Figs 3 and 4). As expected, RERV-H PR is an aspartic protease as shown by structural analysis and inhibition assays (Figs 5 and 6). Due to cleavage by bacterial proteases, we had difficulty in identifying the termini of the mature PR, although a 103 amino acid protein with PR activity was produced (Fig. 4b). From this we conclude that RERV-H does encode a functional protease.

Future work will focus on other aspects of RERV-H function such as Gag assembly and the activity of the other RERV-H enzymes. Preliminary data indicate that the RERV-H Gag polyprotein can assemble into cores when expressed in mammalian cells. The bacterial expression system described here will permit the production of large amounts of normal and altered RERV-H PR proteins, which will be useful for studying the biochemical properties of PR and permit mutational analysis of the proteolytic function. This will be important should RERV-H prove to represent a genuine zoonotic human infection.

ACKNOWLEDGEMENTS

We are grateful to Jan Konvalinka for providing the MPMV PR inhibitor and to Kvido Stříšovský and Dick Campbell for helpful discussion. We thank Christine Shotton for assistance with monoclonal antibody production, John Hackett for providing the polyclonal antibody and Patrick Venables and Robin Weiss for useful discussion and comments on the manuscript. This work was supported by the Arthritis Research Campaign and Abbott Laboratories.

REFERENCES


