Original Paper

Journal of Biomedical Science

J Biomed Sci 1999;6:333–341

Received: February 24, 1999
Accepted: May 5, 1999

Organization of HIV-1 *pol* Is Critical for Pol Polyprotein Processing

Yu-Chung Chang[a] Sung-Liang Yu[a] Wan-Jr Syu[a,b]

[a]Institute of Microbiology and Immunology, and [b]AIDS Prevention and Research Center, National Yang Ming University, Taipei, Taiwan, ROC

Key Words
Autoprocessing · Expression · HIV-1 · Integrase · Protease · Reverse transcriptase

Abstract
The HIV *pol* sequentially encodes protease (PR), reverse transcriptase (RT), and integrase (IN) from the 5′–3′ direction. We explored the significance of this gene arrangement. All six possible gene dispositions were examined. In two situations where *PR* was removed from the leading place and no two genes were in their original location, viral polyprotein processing was abolished. Processing of the polyprotein did not occur when IN was translocated to the front of PR-RT. However, in the following two arrangements, the polyprotein was processed but only at specific sites. First, PR remained in the leading position while the locations of RT and IN were exchanged; viral polyprotein was processed at a site between the upstream transframe peptide (TF) and PR. Second, PR was placed after RT-IN and located at the distal end of Pol. Processing occurred only at the created junction between TF and RT. These results indicated that cleavage after TF occurred autocatalytically but did not proceed to a second site, which needed an extraneous PR for *trans*-action. Therefore, arranging Pol in the order of PR-RT-IN warrants the streamline processing of the polyprotein once the autocleavage is initiated.

Introduction

The human immunodeficiency virus (HIV) is the etiological agent of the acquired immunodeficiency syndrome (AIDS). It belongs to the Retroviridae family. Retroviruses share a basic genomic organization, i.e., *R-U5-gag-pol-env-U3-R*. Structural proteins are expressed by translating an unspliced RNA for the Gag and the GagPol polyproteins. The Env protein is encoded by a single-spliced mRNA. During viral maturation, the polyproteins are cleaved into their respective gene products. The host-cell-encoded proteases mediate the processing of the Env polyprotein, while a viral protease is responsible for the maturation of the viral products from the Gag and GagPol polyproteins [8, 12, 14].

In HIV, simian immunodeficiency virus (SIV), and some other retroviruses, *PR* is within *pol* that sequentially encodes the transframe peptide (termed TF or p6*), protease (PR), reverse transcriptase (RT), and integrase (IN). By an infrequent frameshift mechanism [11], *pol* together with *gag* encodes the GagPol polyprotein where PR is embedded. Presumably, PR is released from the GagPol precursor by autocatalytic cleavage after the virion components are confined in the budding particle. The molecular mechanisms that lead to PR activation, autocatalytic release, and processing of Gag and GagPol are not completely understood.

KARGER
Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1021–7770/99/0065–0333$17.50/0
Accessible online at:
http://BioMedNet.com/karger

W.-J. Syu
Institute of Microbiology and Immunology
National Yang Ming University
Shih-Pai, Taipei 112, Taiwan (ROC)
Tel. +886 2 2826 7112, Fax +886 2 2821 2880, E-Mail wjsyu@ym.edu.tw

Some regulations concerning the maturation and processing of the polyproteins have been examined. Hydrophilic peptides derived from TF were found to inhibit PR activity [16], and deletion of TF enhances Gag polyprotein processing [19, 37]. However, the processing activity of PR with an N-terminal TF extension was obviously reduced when the Gag polyprotein was tested as the protease substrate [27, 38]. These results suggest that positioning of TF in front of PR negatively regulates PR processing.

Regulation of PR processing has also been attributed to other portions of the polyprotein precursor. The disproportionate synthesis of HIV Gag and GagPol polyproteins in the cytoplasm and consequently the Gag-to-Gag and Gag-to-GagPol interactions might attenuate the activation of PR [6]. The P2 segment between the capsid p24 and the nucleocapsid p7 might regulate PR cleavage at a specific site of Gag [20]. Deleting segments of RT or truncating the Pol sequence could impair the processing of Gag and render HIV-1 particles with abnormal morphologies [22].

In the examination of HIV *pol* gene organization, *PR* is sequentially followed by *RT* and *IN*. In other retroviruses such as avian retroviruses, mouse mammary tumor virus, bovine leukemia virus, and human adult T cell leukemia viruses, PR is located at the C-terminal end of the Gag frame and translated together with Gag as a GagPR polyprotein precursor [4, 9, 18, 33]. Experiments have been carried out to test whether HIV-1 PR could be organized in a way similar to that of avian retroviruses. The results indicate that HIV-1 PR placed in-frame at the C-terminus of Gag remained active [13, 37]. In other studies, linking PR to the C-terminus of Vpx or the N-terminus of Vpr [31], or embedding PR into the Nef coding sequence [1] did not prohibit PR activation. Recently, a mutation has been created to prevent the cleavage between PR and RT, and no effect was observed with activities of both PR and RT even though they appeared as a PR-RT fusion protein [3]. All these facts suggest that the enzymatic domains within Pol are relatively independent, and a functional HIV PR may not necessarily be in the natural GagPol context. Therefore, we addressed the question whether the PR activity could be retained after the order of PR, RT, and IN in the Pol polyprotein is reorganized.

Autoactivation and processing of the large GagPol polyprotein could lead to multiple products. Scrambling the domains of Pol could add further complexity to the problem. A recent proviral transfection experiment has revealed that immature PR has less effect on the processing of Pol than that of Gag [27]. Therefore, we used the entire Pol, starting from the 5th codon to the end of the Pol reading frame, as our model precursor polyprotein. Here, we report the effect on autoprocessing after the domains of Pol were rearranged in different orders.

Materials and Methods

Protein Expression

A *Bgl*II/*Nco*I fragment of the HIV-1 provirus HXB2 [23], covering the entire *pol*, was cloned into the T7-RNA polymerase-driven expression vector pRSET-C (Invitrogen, Carlsbad, Calif., USA). With this construct, PR-RT-IN was fused to the C-terminus of a 12-kD peptide, resulting in a conjoint protein of 122 kD. Expression of the fusion protein was induced by adding isopropyl-β-*D*-thiogalactoside (IPTG) [34] to the culture of plasmid-transformed *Escherichia coli* BL21(DE3). Alternatively, *E. coli* BL21 was infected with lambda phage CE6 [24] expressing T7-RNA polymerase that in turn caused overexpression of the targeted gene.

For better target detection, the background proteins of *E. coli* were suppressed by adding rifampicin [25], a bacterial RNA polymerase inhibitor. Plasmid-transformed *E. coli* BL21(DE3) was cultured in Luria-Bertani broth containing 2 μg/ml of *o*-nitrophenyl-β-*D*-fucopyranoside and 100 μg/ml of ampicillin. When the bacterial culture reached an $OD_{600\ nm}$ of 1.0, 1 vol (100 μl) of the bacterial suspension was centrifuged, and the pellet was washed twice with 1 ml of methionine-deficient RPMI (Gibco, Gaithersburg, Md., USA). The pellet was resuspended in the same RPMI and incubated for an additional 100 min at 37°C. Then, IPTG (0.4 m*M*) and rifampicin (200 μg/ml) were added to the culture, and incubation was continued for 1 h. The bacteria were then labeled with 50 μCi/ml of ^{35}S-methionine for 5–15 min. The labeled bacteria were chilled immediately on ice and pelleted promptly. The cell pellets were lysed directly with SDS sample buffer. Proteins were separated by SDS-PAGE and visualized by autoradiography on an X-ray film or a Phosphoimage cassette (Molecular Dynamics, Sunnyvale, Calif., USA).

In some constructs, the HIV proteins were expressed to a level that could be detected with Coomassie blue dye staining after SDS-PAGE separation. Under these circumstances, proteins from unlabeled cells were analyzed using Western blotting [35] with rabbit polyclonal antisera that were specific to PR, RT, and IN, respectively.

To prepare bacterial lysate for the *trans*-cleavage assay or immunoprecipitation, bacteria were lysed as previously described [26]. The lysate was cleared by centrifugation and used directly in the *trans*-cleavage assay (see below). Immunoprecipitation was carried out as previously described [32]. Samples were prepared in 50 m*M* Tris-HCl (pH 7.4) containing 150 m*M* NaCl, 1 m*M* EDTA, 1% Triton X-100, 1% deoxycholate, and 0.1% SDS.

Assembling pol *from DNA Modules*

To generate DNA fragments for gene assembly, proviral DNA HXB2 [23] was used as a template in PCR reactions using Vent (New England Biolabs, Beverly, Mass., USA) or Pfu (Stratagene, La Jolla, Calif., USA) DNA polymerase. The *pol* open reading frame (ORF) was divided into four modules: *TF, PR, RT,* and *IN. TF* and the 3′-flanking region (*FL*) of *pol* were amplified separately with primer pairs M2096/MR2252 and M5093/MR5675 (fig. 1a), respectively. M2096 contains a *Bgl*II site, whereas the antisense primer, MR2252,

incorporates an *Sma*I site and an *Xba*I site. M5093 contains an *Xba*I site, whereas MR5675 includes an *Nco*I site. The PCR-amplified *TF* segment was digested with *Bgl*II and *Xba*I, and that of *FL* was cut with *Xba*I and *Nco*I. These products were three-way ligated with *Bgl*II/*Nco*I-digested pREST-C to generate pTF (fig. 1b). As a result of primer design, the ligated junction between *TF* and *FL* contains an *Sma*I site and an *Xba*I site.

The genes *PR*, *RT*, and *IN* were separately amplified with three different primer pairs to generate the corresponding modules. These primer pairs were: MPR/MRPR for *PR*, MRT/MRRT for *RT*, and MIN/MRIN for *IN* (fig. 1a). The *pol* assembly was then carried out with an approach exemplified by the vector pTF-RPI that encodes Pol in the TF-RT-PR-IN organization (fig. 1b). The *RT* module was digested with *Xba*I and then ligated with the *Sma*I/*Xba*I-digested pTF to generate pTF-R. During this ligation, the 3′-end of *TF* created by the *Sma*I digestion was linked to the 5′-end of the PCR-generated *RT* fragment so that this *Sma*I site was not reconstituted. However, the 3′-end of the *RT* fragment contained an *Sma*I site and an *Xma*I site. They both became unique in the newly generated pTF-R and were available for usage of the next module insertion. With a similar approach, the *PR* module was inserted into pTF-R to form pTF-RP. Then *IN* was inserted into pTF-RP to generate pTF- RPI.

For easy discussion, the *pol*-encoded gene products in the polyproteins are named after the gene organization with three capital letters hereafter. For example, RPI represents the polyprotein derived from pTF-RPI that sequentially encodes reverse transcriptase (R), protease (P), and integrase (I) after TF. RT, PR, and IN are still reserved to denote the reverse transcriptase, the protease, and the integrase, respectively, and they are discernible from the corresponding genes *RT*, *PR*, and *IN* that are italicized.

Plasmids for all possible Pol organizations including pTF-R*IP were constructed that way except for pTF-PRI and pTF- RIP. DNAs were sequenced for all the ligated junctions. The pTF-PRI and pTF-RIP vectors were generated by amplifying the DNA segment covering both *RT* and *IN* with the MRT/MRIN primer pair. This two-unit DNA segment was then treated with *Xba*I and inserted into appropriate plasmids to create the designated constructs. As exemplified in the construction of pTF-RIP, this *Xba*I-treated RT-IN module was inserted into the *Sma*I/*Xba*I-digested pTF (fig. 1b) to form pTF-RI. The *PR* module was then inserted into pTF-RI to form pTF-RIP according to the strategy outlined in figure 1b. Consequently, the DNA sequence between *RT* and *IN* in pTF-RIP was retained in the natural context. In contrast, the pTF-R*IP has an extra Pro inserted between RT and IN as a result of stepwise single-unit DNA manipulation.

Since *pol* was reorganized in different gene orders, the junctions between adjacent genes were changed accordingly. The amino acid sequences constituting these sites were lined up with those of the natural junctions of PRI (fig. 2). Residues were designated relative to the scissile bond. P1 and P1′ are the first amino acids upstream and downstream of the scissile bond, respectively. Some of these sequences deviated from the preferred PR cleavage sequences due to DNA manipulation. However, at least one junction with a sequence demonstrated to be PR sensitive was coined in each of the Pol constructs (fig. 2).

A second set of the same constructs was prepared with a *PR* module encoding an inactive HIV-1 protease. The inactive *PR* [5] was created by site-directed mutagenesis of replacing Leu_{23} and Asp_{25} with Ile and Asn, respectively. The mutated nucleotides were confirmed by DNA sequencing.

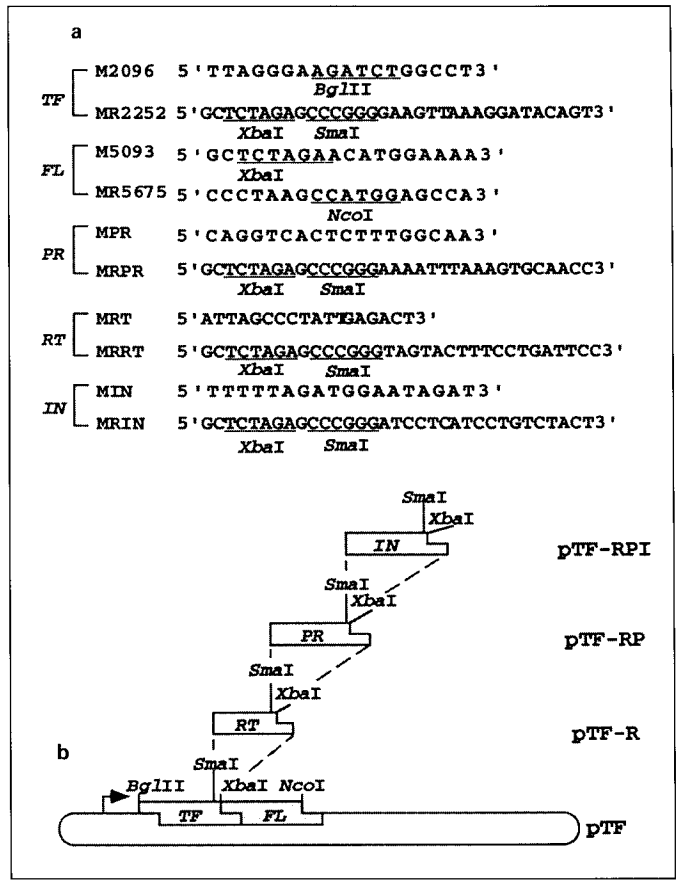

Fig. 1. Strategy to construct expression vectors for reorganizing Pol polyproteins. **a** Primer pairs used to amplify target DNA from HXB2 and the incorporated restriction enzyme sites. **b** Sequential insertions of the amplified DNA modules into the expression vector pRSET-C. Blunt or sticky ends of joined DNA fragments are represented by different shapes, and the sites recognized by appropriate restriction enzymes are labeled. The ORF of the assembled plasmid initiates at a site indicated with a bent arrow and ends at the point defined by the 3′-end of the *Xba*I site.

Trans-*Cleavage Assay*

To examine whether a particular Pol polyprotein junction could be cleaved by HIV-1 PR in a *trans* manner, a protease lysate was added directly to the Pol-containing lysate without further purification [34]. After incubation at 37°C for 60 min, proteolysis of the polyprotein was followed by Western blotting using product-specific antibodies.

Purification of His-Tagged Fusion Proteins

The bacteria were cultured as previously described [2], except that the bacterial pellet was resuspended and lysed in 20 m*M* Tris-HCl (pH 8.0) containing 6 *M* guanidine-HCl and 0.5 *M* NaCl. The mixture was shaken vigorously at room temperature for 1 h. Cell debris was removed by centrifugation at 4°C. His-tagged proteins were purified as described [10] with a slight modification. In brief, the supernatant was loaded onto a nickel ion column (Probond™;

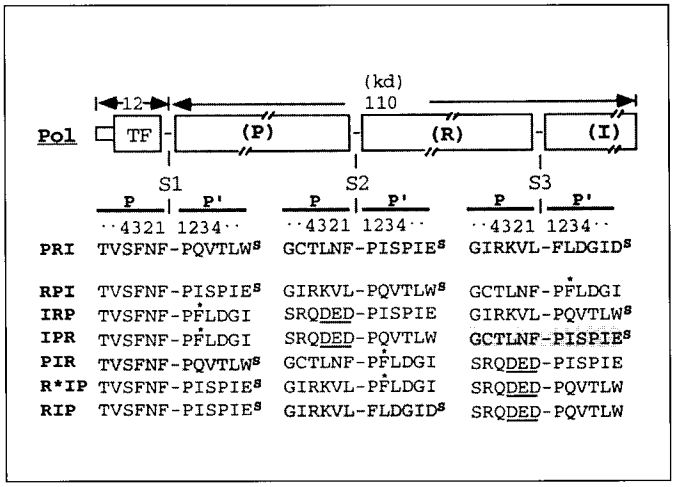

Fig. 2. Summary of the reorganized *pol* ORF. The viral genes encoded by *pol* were reorganized in different orders and cloned into the pRSET-C expression vector as illustrated in figure 1. The encoded protein modules assembled from DNA blocks are represented by different open boxes. Their assembly junctions are denoted S1, S2, and S3. The amino acid sequences in these junctions are lined up with those present in the natural junctions of PRI. P and P′ denote the residue positions relative to the scissile peptide bond (P1 and P1′ are the first amino acids upstream and downstream of the scissile bond, respectively). Sequences identical to a natural junction are shadowed. Two types of junctions are resistant to PR: one is that with a Phe residue (asterisk) in the P2′ position, and the other is that with three consecutive acidic residues (underlined) in P1–P3 positions. The rest of the sequences demonstrated to be sensitive to PR are labeled with a superscript 'S' (see 'Discussion').

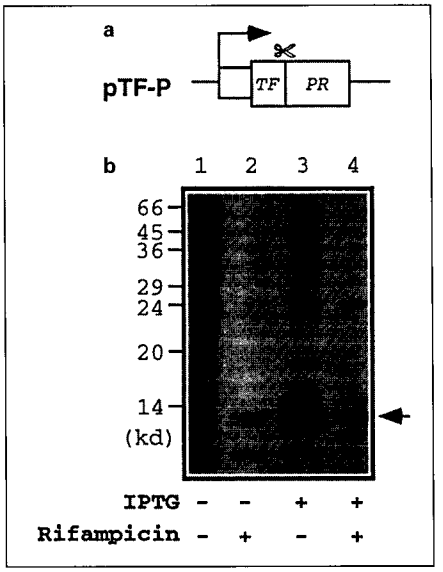

Fig. 3. Detection of low-level expression and processing of HIV-1 protease using metabolic labeling. The fusion protein encoded by the plasmid pTF-P was metabolically labeled with ^{35}S-Met, in the presence or absence of IPTG and rifampicin. Cell lysates were separated by SDS-PAGE, and isotope-labeled proteins were visualized by autoradiography. **a** ORF encoded by the plasmid pTF-P. Translation starts at the site indicated by a bent arrow. The protease-processing site of the expressed fusion protein is denoted with scissors. The processing products are similar in size: TF-leading peptide: 12 kD; PR: 11 kD. **b** Autoradiogram of cell lysates labeled in the presence or absence of IPTG and rifampicin. The arrow labels the unresolved products of the autoprocessed TF-PR fusion protein.

Qiagen, Valencia, Calif., USA) that was pre-equilibrated with the lysis buffer. After loading, the Ni^{2+} column was washed with 20 m*M* Tris-HCl (pH 8.0) containing 8 *M* urea, 0.5 *M* NaCl, and 20 m*M* imidazole. The proteins bound on the column were refolded with 50 ml TBS buffer (20 m*M* Tris-HCl, pH 8.0, 0.5 *M* NaCl) and eluted off the column using TBS containing 50 m*M* EDTA. The protein solution was concentrated in dialysis tubing by embedding in PEG-6000 powder.

Results

Establishment of the Model System

Protein expression and polyprotein processing in bacteria transformed with pTF-PRI were examined. This plasmid encodes a full-length Pol polyprotein in the native configuration, and the encoded polyprotein is expected to be processed as reported [15]. We could not detect any overexpressed proteins in the cell lysate by SDS-PAGE analysis and Coomassie blue dye staining (data not shown). Probably, the recombinant proteins were expressed at a level below the detection limit of the dye staining method. To circumvent this difficulty, we labeled the induced proteins with ^{35}S-Met and, at the same time, suppressed the synthesis of host proteins by adding rifampicin to the culture medium.

The feasibility of this approach was assessed on cells containing pTF-P in which the *PR* gene module alone was ligated into pTF (fig. 3a). In this plasmid, the encoded precursor has a size of 23 kD, and autoprocessing between TF and PR would give 12- and 11-kD products that are not resolvable on SDS-PAGE. As shown in figure 3b, metabolic labeling of the pTF-P-transformed bacteria without any treatment yielded many labeled protein bands (lane 1). The addition of rifampicin to the culture significantly suppressed the isotopic labeling of host proteins, and a weak 11- to 12 kD band could be detected (lane 2). When IPTG alone was added to the culture, the specific band was induced significantly (lane 3). However, the best

result was achieved by adding IPTG and rifampicin simultaneously (lane 4).

The *pol* ORF may initiate internal translation during protein synthesis [36], and the translated products may encounter nonspecific degradations. To identify these nonspecific products, equivalent constructs were prepared with an inactive PR and included in the experiments. For clarity, polyproteins containing an inactive PR were denoted with '℗', as in the construct of ℗RI.

As exemplified by the expression patterns of pTF-PRI and pTF-℗RI, the processing of the wild-type PRI polyprotein should give the 12-kD TF-leading peptide, the 66-kD RT product, the 31-kD IN, and the 11-kD PR (fig. 4a). The 66-kD RT product may further mature into 66- and 51-kD heterodimers. As shown in figure 4b (lane 1), the 66- and the 51-kD RT products and the 31-kD IN protein can be visualized, while the small products have run off that particular gel. The identities of the 66- and 51-kD protein bands were confirmed by immunoprecipitation with anti-RT antibodies and denaturing gel electrophoresis (fig. 4c). In contrast, the SDS-PAGE patterns of cells harboring pTF-℗RI contained many protein bands, the largest of which was 122 kD in size (fig. 4b, lane 2). Since PR was inactive in ℗RI and no specific processing could occur, the largest product probably represented the unprocessed polyprotein. Indeed, this notion was supported by the fact that the 122-kD protein can be precipitated by anti-RT, anti-IN, or anti-PR antibodies, but not by a non-immune serum (fig. 4d).

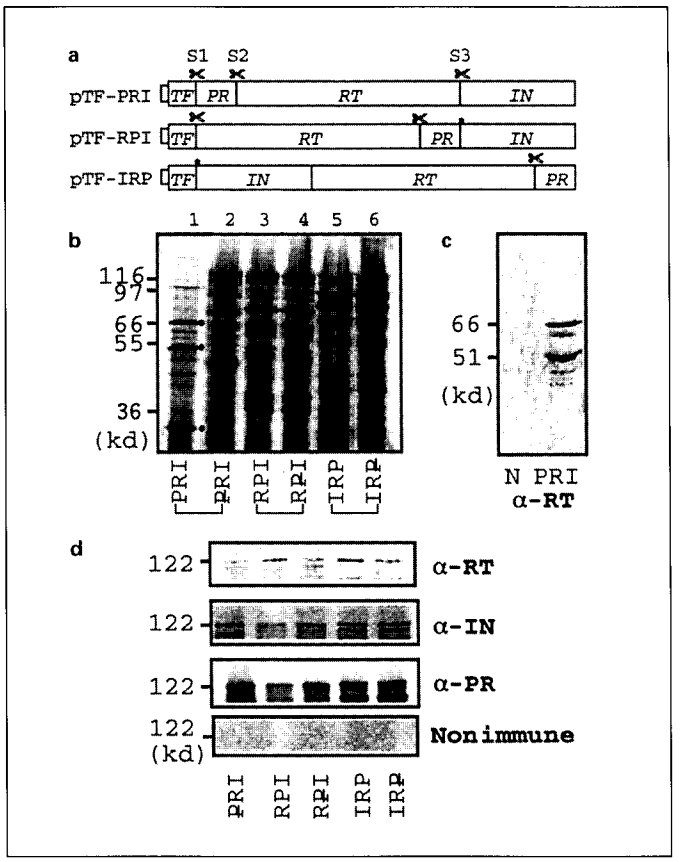

Fig. 4. Expression patterns of the polyproteins PRI, RPI, and IRP. **a** Organization of the ORF encoded by the corresponding expression vectors. Junctions with Phe at the P2′ position are attested with asterisks; junctions with cleavage observed in either autoprocessing or *trans*-cleavage assay are marked with scissors. **b** Protein expression profiles of individual Pol constructs. Proteins synthesized in the plasmid-transformed bacteria were labeled with ^{35}S-Met in the presence of both IPTG and rifampicin. Total bacterial lysates were separated by SDS-PAGE and visualized by autoradiography. Dots label the 66- and 51-kD RT products and the 31-kD IN protein, respectively. Constructs containing an inactive version of the PR are denoted with a ℗. **c** Autoradiogram of the electrophoretically separated RT products processed from PRI. The products were obtained by immunoprecipitation using anti-RT antibodies. N = Control experiment using the expression lysate derived from pTF (fig. 1b). **d** The 122-kD products in **b** were confirmed as the polyprotein precursors containing PR, RT, and IN by immunoprecipitation using anti-RT (α-RT), anti-IN (α-IN), anti-PR (α-PR), and nonimmune antibodies, respectively. The analyses were similar to those done in **c**.

Expression and Processing of the RPI and IRP Polyproteins

The question of whether autoprocessing could occur if the three genes in *pol* were completely scrambled was then addressed. The criteria for complete scrambling are: (1) no two genes are in their natural position, and (2) *PR* is not proximal to *TF*. Constructs in the plasmids pTF-RPI and pTF-IRP (fig. 4a) met these criteria and were examined. Constructs in pTF-R℗I and pTF-IR℗ encompassing inactive PR were included in these experiments for comparison. The isotope-labeled protein profiles derived from RPI and R℗I (fig. 4b, lanes 3, 4) were indistinguishable. The largest translational products have a size close to 122 kD and were precipitable by antibodies specific to individual protein components (fig. 4d). They probably represent the unprocessed precursors. Proteins smaller than the 122-kD polyprotein were detected both in RPI and R℗I samples. They have the same electrophoretic pattern on denaturing gels and are presumably products unrelated to processing. Thus, RPI was not processed to a detectable level even though the construct contains scissible junctions (fig. 2). Similar observations were found with the analysis of IRP and IR℗ (fig. 4b, lanes 5, 6). Therefore, complete scrambling of the *pol* organization abolished the polyprotein processing.

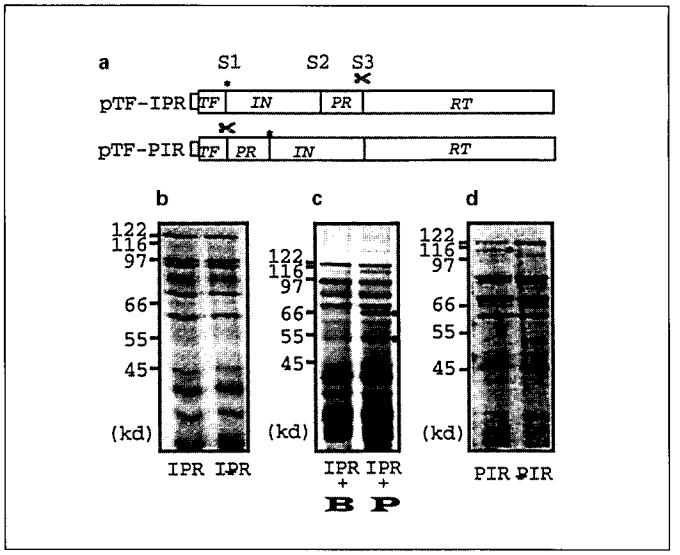

Fig. 5. Expression and processing of the polyproteins IPR and PIR. **a** ORF and gene organizations. Labels are identical to those in figure 4. **b** Comparison of expression patterns between IPR and I$\overset{\bullet}{P}$R. Metabolic labeling and SDS-PAGE analysis of polyproteins were carried out as in figure 4b. **c** The ^{35}S-Met-labeled IPR was processed to yield the 66- and 51-kD RT dimeric products by adding lysate containing the HIV-1 PR. Analysis was carried out by using SDS-PAGE and autoradiography. **d** Comparison of expression patterns between PIR and $\overset{\bullet}{P}$IR. Products from specific processing are designated with dots. P = PR lysate; B = control lysate without PR.

Expression and Processing of the IPR Polyprotein

We then examined constructs where Pol was partially disorganized. In the polyprotein derived from pTF-IPR, IN was translocated proximal to TF (fig. 5a), and the natural junction between PR and RT was preserved. Comparison of the protein profiles expressing polyproteins IPR and I$\overset{\bullet}{P}$R on SDS-PAGE (fig. 5b), again, indicated that no detectable processing had occurred in the polyprotein.

The S3 junction between PR and RT in the IPR polyprotein (fig. 5a) retains the natural context. It should be cleaved by the *trans*-action of a mature PR. To test this hypothesis, IPR-containing lysates were supplemented with an active PR, and the reaction yielded extra bands on denaturing gel at 66 and 51 kD (fig. 5c). The estimated sizes of these products are consistent with those expected for the heterodimeric RT. They were specifically and reproducibly generated by the addition of the PR lysate, not by the control lysate without PR (fig. 5c). The identities of these two bands were substantiated by immunoprecipitation using anti-RT antibodies (data not shown). Thus, IPR cannot mediate autocleavage at the S3 site, and the cleavage apparently needs an extraneous PR.

Expression and Processing of the PIR Polyprotein

In the PIR polyprotein, PR was retained in the original position while the order of RT and IN was reversed (fig. 5a). As a result, only S1 retains the natural sequence (fig. 2). A cleavage at S1 would generate a 110-kD product. The expression profiles of PIR and $\overset{\bullet}{P}$IR are shown in figure 5d, indicating that PIR was indeed autoprocessed to the 110-kD product. However, some 122-kD precursor still remained unprocessed, suggesting that autoprocessing is inefficient.

Expression and Processing of the RIP Polyprotein

In the RIP polyprotein, PR was translocated to the end of the polyprotein, and RT was linked to IN so that the S2 site retained its natural context (fig. 2). However, the S1 site joining TF and RT may also be PR sensitive since it has no apparent conflict with the preferred PR cleavage sequences [7, 21, 28]. Therefore, the processing at S2 may be complicated by the products of S1 cleavage. This issue was circumvented by taking advantage of the R*IP construct. The S2 and S3 on R*IP are presumably resistant to PR cleavage, due to the presence of Phe [17] at the P2′ position of S2 and serious sequence deterioration at S3 (see Discussion). As shown in figure 6b (lanes 2, 3), a 110-kD protein appeared in the cell lysate containing R*IP but not in that containing R*I$\overset{\bullet}{P}$. The result suggests that the S1 junction between TF and RT was autoprocessed. Again, this autoprocessing was not highly efficient as evidenced by the amount of residual 122-kD precursor. No additional PR proteolytic products were observed, indicating that S2 and S3 were not processed as expected.

We further used the pTF-R plasmid (fig. 6a) to demonstrate that the created S1 between TF and RT is readily digestible by HIV-1 PR. The pTF-R construct has RT alone linked to the C-terminus of TF, and their junction is identical to that of S1 in R*IP. With this vector, a 78-kD recombinant protein was overexpressed and could be visualized on SDS-PAGE with Coomassie blue dye staining. The recombinant protein was incubated with HIV-1 PR, and the products were analyzed by Western blotting using anti-RT antibodies. Typical results are shown in figure 6c. In the presence of a lysate containing the active PR, the 78-kD recombinant protein (lane 3) was efficiently processed into the expected 66- and 51-kD RT products (lane 1). These RT products were not generated by a similar lysate without PR (lane 2).

The above evidence strongly suggests that the cleavage of S1 in R*IP was mediated by PR and the embedded PR in R*IP must be activated. We then asked whether the natural S2 junction in RIP could be processed autocatalytically. The expression profile of RIP was then compared to that of R*IP (fig. 6b, lanes 2, 4). Neither an adverse decrease of the 110-kD intermediate nor a significant increase of the 66- and 51-kD RT products (compared to lane 1) was recognized. This result suggests that the activated PR in the context of RIP was autoprocessed at the S1 site only whereas processing at the second (S2) site was negligible.

To ascertain that the S2 junction is cleavable but not processed well in RIP, the expressed products of pTF-RIP were further incubated with a lysate containing the active PR. The pTF-RI plasmid that encodes a 110-kD TF-RT-IN polyprotein, abbreviated as RI here (fig. 6a), was also included in the experiments to facilitate the analysis. The S1 and S2 sites in the RI polyprotein are equivalent to those in the RIP polyprotein. *Trans*-action of PR on RI would generate RT and IN as products. Analogously, processing of RIP at S1 and S2 would also give RT products and possibly an IN-PR fusion product (fig. 6a). The RT products processed from both RI and RIP should have the same sizes and be easily identified on gel. Indeed, the 66- and 51-kD RT products were both evident after adding the active HIV-1 PR to the expressed RI (fig. 6d, lanes 1, 2) and RIP (fig. 6d, lanes 3, 4). Therefore, it was concluded that PR in the RIP polyprotein was activated and mediated the cleavage at the S1 junction. However, the S2 site was not disjointed although it is cleavable.

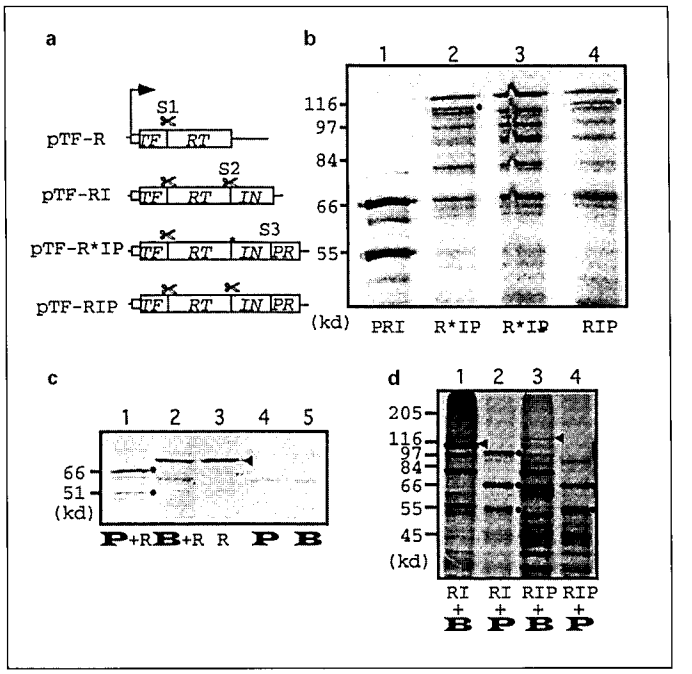

Fig. 6. Expression and processing of the polyproteins R*IP and RIP. **a** ORF and gene organizations. **b** Expression patterns of polyproteins analyzed with metabolic labeling and SDS-PAGE. See text and figure 2 for the difference between RIP and R*IP. **c** *Trans*-cleavage at S1 between TF and RT. Nickel-column-purified R (product of pTF-R) was incubated with an *E. coli* lysate containing PR or a control lysate without PR as indicated. Incubation products were analyzed with Western blotting using rabbit anti-RT antibodies. **d** *Trans*-cleavage at S2 between RT and IN in RI (product of pTF-RI) and RIP. ^{35}S-labeled proteins were treated with or without PR as in **c** and analyzed with SDS-PAGE and autoradiography. The precursor polyproteins are marked with arrowheads. Products from specific processing are labeled with dots. The 66- and 51-kD bands were presumably the RT products obtained by processing of RI and RIP. The 98-kD product detected in the proteolysis of RI was probably a processing intermediate that resulted from the cleavage at S1. R, RI = Products of pTF-R and pTF-RI; P = PR lysate; B = control lysate without PR.

Discussion

We have scrambled the arrangement of *PR*, *RT*, and *IN* in the HIV-1 *pol* and constructed all possible Pol polyproteins in different organizations. Examination of the processed products after gene expression revealed that the naturally evolved Pol configuration has the highest polyprotein-processing efficiency at every cleavage site. The Pol-processing efficiencies in the other five configurations were either reduced or became undetectable.

The processing of the wild-type Pol is highly efficient [15], and few intermediate products can be discerned. Inefficient processing of the polyproteins PIR and RIP allows the processing intermediates to be detected. In PIR, processing occurs at the S1 site joining TF and PR (fig. 5d). Similarly, processing of the RIP polyprotein also occurs at the S1 site (fig. 6). Even though the S2 site between RT and IN has a natural cleavage sequence, it remains uncut. This fact suggests that TF contains the information for signaling the first cleavage at its C-terminal junction even though it has a negative effect in attenuating the PR activity [16, 19, 27, 37].

A model for polyprotein processing has been proposed [30] based on studies on a mini-Pol precursor protein (PR linked to short flanking sequences). The model depicts the polyprotein dimerization before PR is activated. The activated PR then cleaves the junction between TF and PR intramolecularly to release the partially processed PR. The cleavage at the second junction between PR and RT

may subsequently occur intermolecularly [30]. In our observation with the reorganized Pol (i.e., PIR and RIP), autoprocessing at S1 is consistent with the proposed model based on autocleavage of the mini-Pol precursor [30]. After the S1 cleavage, the Pol remaining part has to undergo further processing at other cleavable sites. The S2 site on RIP being in the natural context is an obvious candidate. However, cleavage at this junction did not occur unless extraneous PR was added (fig. 6d). Therefore, availability of cleavable site(s) does not warrant the streamline processing of Pol. After the initiation of autocleavage, the efficient release of a mature PR from the polyprotein probably is a critical step in the processing cascade [29].

Three scrambled Pol polyproteins lose the activity of autoprocessing. The simplest explanation is that the domains in Pol are not completely independent so that the domain scrambling may restrict proper folding, PR activation, or autoprocessing of the precursor. An alternative possibility is that all junctions are blocked for cleavage. However, this possibility could be excluded by examining the peptide sequences of the domain junctions. The junctions shown in figure 2 can be categorized into five groups, and at least one PR-sensitive site is contained in each of the constructs. In group 1, the junctions are preserved in a natural sequence. In group 2, the S1 junctions join TF and RT in the polyproteins RPI, R*IP, and RIP. They are PR sensitive and have been experimentally proven in the R*IP and RIP constructs (fig. 6). Group 3 has the junction connecting RT and PR at S2 of RPI and S3 of IRP. This junction has a composite sequence derived from two scissile sites. They are expected to be PR sensitive, and, indeed, they were PR cleavable as observed in a *trans*-cleavage assay [unpubl. data]. Group 4 consists of the S1 junctions of IRP and IPR, the S2 junctions of PIR and R*IP and the S3 junction of RPI. They all have Phe at the P2′ position and reportedly are resistant to PR cleavage [17]. Our observation that no cleavage occurred at the S2 junction of R*IP (fig. 6) [unpubl. data] supports this notion. Group 5 consists of the junctions connecting IN to PR or to RT, and they include S2 of IRP or IPR and S3 of PIR, R*IP, or RIP. These junctions have three consecutive acidic residues in the P1–P3 positions. These sequences vastly differ from the PR-cleaved structures that frequently contain hydrophobic residues at the P1 position [7, 21, 28]. They are not expected to be cleaved by PR. In a preliminary experiment using TF-IN-RT polyprotein in a *trans*-cleavage assay, the IN-RT junction indeed was not cleavable (data not shown).

The group 4 junctions could be modified to delete the P2′ Phe, as in the cases of R*IP and RIP, so that these sequences could be much like a PR-preferred sequence. We did not do all the changes because no enhancement on autoprocessing was observed when R*IP and RIP were compared (fig. 6b). The group 5 junctions consist of the C-terminus of IN that ends with three consecutive amino acids in HIV-1. This C-terminal end is not conserved in related viruses such as HIV-2 and SIV, and, therefore, the sequence may be modified without adverse effects on the IN domain. We did not carry out the modification for the following reasons: (1) our reorganized Pol constructs all contain at least one junction sensitive to extraneous PR, and (2) as observed experimentally, the scrambled Pol constructs (i.e., RPI and RIP) with two PR-cleavable junctions did not appear to autoprocess better than those with a single susceptible site (e.g. PIR and R*IP).

In summary, our results indicate that proper organization of Pol is critical for the autoprocessing of the polyprotein precursor. Scrambling the order of functional domains in Pol leads to a decreasing efficiency or complete loss of autoprocessing activity at susceptible sites. While linking TF to PR negatively regulates the activity of PR [19, 27, 38], Cherry et al. [3] recently mutated the junction between PR and RT and demonstrated that PR-RT fusion does not interfere with the PR-processing activity. These facts suggest that the PR domain in the natural Pol context may interact with TF but is rather independent of RT. In our construct of PIR, the TF-PR neighborhood relationship was preserved, and the C-terminus of PR was similarly fused to IN-RT rather than RT alone. The PR activity was retained, while the processing efficiency was obviously reduced (fig. 5d). Therefore, HIV Pol must be organized in the order of PR-RT-IN to warrant an appropriate conformation, which probably has the least domain constraint, to facilitate streamline processing of the polyprotein.

Acknowledgments

We thank M.F. Tam for critical reading of the manuscript. This work was supported in part by grant NSC86-2314-B-010-037 from the National Science Council and an award from the Medical Research Advancement Foundation in memory of Dr. Chi-Shuen Tsou.

References

1 Arrigo SJ, Haines JK, Huffman KM. Intrinsic activity of human immunodeficiency virus type 1 protease heterologous fusion proteins in mammalian cells. DNA Cell Biol 14:15–23; 1995.

2 Chang YC, Ching TT, Syu WJ. Assaying the activity of HIV-1 integrase with DNA-coated plates. J Virol Methods 59:135–140;1996.

3 Cherry E, Liang C, Rong L, Quan Y, Inouye P, Li X, Morin N, Kotler M, Wainberg MA. Characterization of human immunodeficiency virus type-1 (HIV-1) particles that express protease-reverse transcriptase fusion proteins. J Mol Biol 284:43–56;1998.

4 Craven RC, Bennett RP, Wills JW. Role of the avian retroviral protease in the activation of reverse transcriptase during virion assembly. J Virol 65:6205–6217;1991.

5 Darke PL, Leu CT, Davis LJ, Heimbach JC, Diehl RE, Hill WS, Dixon RA, Sigal IS. Human immunodeficiency virus protease: Bacterial expression and characterization of the purified aspartic protease. J Biol Chem 264:2307–2319;1989.

6 Gatlin J, Arrigo SJ, Schmidt MG. HIV-1 protease regulation: The role of the major homology region and adjacent C-terminal capsid sequences. J Biomed Sci 5:305–308;1998.

7 Griffiths JT, Phylip LH, Konvalinka J, Strop P, Gustchina A, Wlodawer A, Davenport RJ, Briggs R, Dunn BM, Kay J. Different requirements for productive interaction between the active site of HIV-1 proteinase and substrates containing -hydrophobic*hydrophobic- or -aromatic*pro-cleavage sites. Biochemistry 31: 5193–5200;1992.

8 Hallenberger S, Bosch V, Angliker H, Shaw E, Klenk H, Garten W. Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. Nature 360:358–361;1992.

9 Hizi A, Henderson LE, Copeland TD, Sowder RC, Hixson CV, Oroszlan S. Characterization of mouse mammary tumor virus *gag-pro* gene products and ribosomal frameshift site by protein sequencing. Proc Natl Acad Sci USA 84: 7041–7045;1987.

10 Holzinger A, Phillips KS, Weaver TE. Single-step purification/solubilization of recombinant proteins: Application to surfactant protein B. Biotechniques 20:804–808;1996.

11 Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. Characterization of ribosomal frame shifting in HIV-1 *gag-pol* expression. Nature 331:280–283;1988.

12 Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RA, Scolnick EM, Sigal IS. Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci USA 85:4686–4690;1988.

13 Kotler M, Arad G, Hughes SH. Human immunodeficiency virus type 1 *gag*-protease fusion proteins are enzymatically active. J Virol 66: 6781–6783;1992.

14 Kramer RA, Schaber MD, Skalka AM, Ganguly K, Wong-Staal F, Reddy EP. HTLV-III gag protein is processed in yeast cells by the virus pol-protease. Science 231:1580–1584;1986.

15 Loeb DD, Hutchison CA III, Edgell MH, Farmerie WG, Swanstrom R. Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases. J Virol 63:111–121;1989.

16 Louis JM, Dyda F, Nashed NT, Kimmel AR, Davies DR. Hydrophilic peptides derived from the transframe region of Gag-Pol inhibit the HIV-1 protease. Biochemistry 37:2105–2110; 1998.

17 Margolin N, Heath W, Osborne E, Lai M, Vlahos C. Substitutions at the P2′ site of gag p17-p24 affect cleavage efficiency by HIV-1 protease. Biochem Biophys Res Commun 167: 554–560;1993.

18 Nam SH, Kidokoro M, Shida H, Hatanaka M. Processing of gag precursor polyprotein of human T-cell leukemia virus type I by virus-encoded protease. J Virol 62:3718–3728;1988.

19 Partin K, Zybarth G, Ehrlich L, DeCrombrugghe M, Wimmer E, Carter C. Deletion of sequences upstream of the proteinase improves the proteolytic processing of human immunodeficiency virus type 1. Proc Natl Acad Sci USA 88:4776–4780;1991.

20 Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet PV, Klein CA, Swanstrom R. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. J Virol 68:8017–8027;1994.

21 Pettit SC, Simsic J, Loeb DD, Everitt L, Hutchison CA III, Swanstrom R. Analysis of retroviral protease cleavage sites reveals two types of cleavage sites and the structural requirements of the P1 amino acid. J Biol Chem 266: 14539–14547;1991.

22 Quillent C, Borman AM, Paulous S, Dauguet C, Clavel F. Extensive regions of Pol are required for efficient human immunodeficiency virus polyprotein processing and particle maturation. Virology 219:29–36;1996.

23 Ratner L, Fisher A, Jagodzinski LL, Mitsuya H, Liou RS, Gallo RC, Wong-Staal F. Complete nucleotide sequences of functional clones of the AIDS virus. AIDS Res Hum Retroviruses 3:57–69;1987.

24 Rosenberg AH, Lade BN, Chui DS, Lin SW, Dunn JJ, Studier FW. Vector for selective expression of cloned DNAs by T7 RNA polymerase. Gene 56:125–135;1987.

25 Tarbor S, Richardson CC. A bacteriophage T7 DNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc Natl Acad Sci USA 82:1074–1078;1985.

26 Tatsumi H, Kajiyama N, Nakano E. Molecular cloning and expression in *Escherichia coli* of a cDNA clone encoding luciferase of a firefly, *Luciola lateralis*. Biochim Biophys Acta 1131: 161–165;1992.

27 Tessmer U, Krausslich H. Cleavage of human immunodeficiency virus type 1 proteinase from the N-terminally adjacent p6* protein is essential for efficient Gag polyprotein processing and viral infectivity. J Virol 72:3459–3463; 1998.

28 Tozser J, Weber IT, Gustchina A, Blaha I, Copeland TD, Louis JM, Oroszlan S. Kinetic and modeling studies of S3-S3′ subsites of HIV proteinases. Biochemistry 31:4793–4800; 1992.

29 Vogt VM. Proteolytic processing and particle maturation. Curr Top Microbiol Immunol 214: 95–132;1996.

30 Wondrak EM, Nashed NT, Haber MT, Jerina DM, Louis JM. A transient precursor of the HIV-1 protease. Isolation, characterization, and kinetics of maturation. J Biol Chem 271: 4477–4481;1996.

31 Wu X, Liu H, Xiao H, Kappes JC. Proteolytic activity of human immunodeficiency virus Vpr- and Vpx-protease fusion proteins. Virology 219:307–313;1996.

32 Yan BS, Lee KM, Liu SH, Syu WJ. Characterization of monoclonal antibodies to the 26-kDa glutathione S-transferase of *Schistosoma japonicum*. Hybridoma 15:429–433;1996.

33 Yoshinaka Y, Katoh I, Copeland TD, Smythers GW, Oroszlan S. Bovine leukemia virus protease. Purification, chemical analysis, and in vitro processing of gag polyproteins. J Virol 57: 826–832;1986.

34 Yu SL, Wang N, Liou CY, Syu WJ. Assay of HIV-1 protease activity by use of crude preparations of enzyme and biotinylated substrate. J Virol Methods 53:63–73;1995.

35 Yu SL, Chou MJ, Tam MF, Lee TH, Syu WJ. Expression and antigenicity of human immunodeficiency virus type-1 transmembrane gp41 in insect cells. Biochem Biophys Res Commun 191:207–213;1993.

36 Zervos P, Hassell T, Van Frank R, Lai MT. Characterization of an internally initiated integrase protein of HIV-1 produced in *E. coli*. Biochem Biophys Res Commun 170:1061–1066; 1990.

37 Zybarth G, Carter C. Domains upstream of the protease (PR) in human immunodeficiency virus type 1 Gag-Pol influence PR autoprocessing. J Virol 69:3878–3884;1995.

38 Zybarth G, Krausslich H, Partin K, Carter C. Proteolytic activity of novel human immunodeficiency virus type 1 proteinase proteins from a precursor with a blocking mutation at the N terminus of the PR domain. J Virol 68: 240–250;1994.