

Proteolytic Processing of Ty3 Proteins Is Required for Transposition

JACQUELINE KIRCHNER AND SUZANNE SANDMEYER*

Department of Microbiology and Molecular Genetics, University of California, Irvine, Irvine, California 92717-4025

Received 6 July 1992/Accepted 22 September 1992

Ty3 is a retroviruslike element found in *Saccharomyces cerevisiae*. It encodes *GAG3* and *GAG3-POL3* polyproteins which are processed into mature proteins found in the Ty3 viruslike particle. In this study, the region encoding a protease that is homologous to retroviral aspartyl proteases was identified and shown to be required for production of mature Ty3 proteins and transposition. The Ty3 protease has the Asp-Ser-Gly consensus sequence found in *copia*, Ty1, and Rous sarcoma virus proteases, rather than the Asp-Thr-Gly found in most retroviral proteases. The Asp-Ser-Gly consensus is flanked by residues similar to those which flank the active sites of cellular aspartyl proteases. Mutations were made in the Ty3 active-site sequence to examine the role of the protease in Ty3 particle maturation and to test the functional significance of the Ser active-site variant in the consensus sequence. Mutation of the active-site Asp blocked processing of Gag3 and Gag3-Pol3 and allowed identification of a *GAG3-POL3* polyprotein. This protein was turned over rapidly in cells expressing the mutant Ty3. Changing the active-site Ser to Thr caused only a modest reduction in the levels of certain Ty3 proteins. Five putative cleavage sites of this protease in Ty3 *GAG3* and *GAG3-POL3* polyproteins were defined by amino-terminal sequence analysis. The existence of an additional protein(s) of unknown function, encoded downstream of the protease-coding region, was deduced from the positions of these amino termini and the sizes of known Ty3 proteins. Although Ty3 protease cleavage sites do not correspond exactly to known retroviral protease cleavage sites, there are similarities. Residues P3 through P2' in the regions encompassing each of the five sites are uncharged, and no P1 position is occupied by an amino acid with a branched beta carbon.

The retroviral core contains an aspartyl protease (PR) of 12 to 15 kDa in mass (for a review, see references 42 and 53). Alignment of the protease sequences of retroviral and cellular aspartyl proteases shows two regions of conserved sequence: Asp-Thr/Ser-Gly in the amino-terminal portion of each protease domain and H-H-Gly (where H is Ile, Leu, Trp, or Val) in the carboxyl-terminal portion of each domain (45, 57). In avian sarcoma/leukosis virus, PR is encoded at the downstream end of *gag*, while in some mammalian retroviruses, it is contained in a separate reading frame between *gag* and *pol*, and in others, the coding region overlaps with *gag* but is primarily contained within the upstream end of *pol*. Pulse-labeling studies have shown that both structural and catalytic proteins are synthesized as polyprotein precursors which are found in immature virions. Maturation of the virus particles coincides temporally with the appearance of processed proteins. Mutagenesis of the PR coding domain can block particle maturation and processing, implicating PR in the maturation process (9, 10, 20, 23, 46, 47, 55, 56). More recently, retroviral proteases have been shown to be active on peptide substrates simulating polyprotein cleavage sites (24–26) and in heterologous systems such as *Saccharomyces cerevisiae* (27) or *Escherichia coli* (11, 26), when produced as part of *pol*, thereby confirming that PR is the processing enzyme.

Solution of the X-ray crystal structure for both the Rous sarcoma virus (40) and human immunodeficiency virus (41, 59) PR, information from mutagenesis of several retroviral PR species (20, 23, 33–35, 55), and studies with variant substrates have greatly expanded our understanding of both

the structure and the function of this enzyme. The similarity of PR to cellular aspartyl proteases twice its mass prompted Pearl and Taylor to propose that PR forms a dimer resembling the cellular aspartyl proteases within which a PR domain is repeated (45), and this has been confirmed by structural studies (19, 24, 38, 40, 41). More recently, synthetic dimers have been shown to have intracellular activity in the absence of particle formation, as predicted if dimerization is the rate-limiting step of retroviral protease activation (5, 28). Comparison of synthetic and natural substrates for retroviral proteases has shown that there is not a strict cleavage site consensus (24, 25, 51). Natural cleavage sites are within hydrophobic regions, especially from P2 to P2', and do not contain beta-branched amino acids at position P1. Some cleavage sites are only recognized in the native context, arguing that higher-order structures also play a role in cleavage site recognition (34, 58). The hydrophobic context, together with synthetic peptide substrates and the X-ray crystal structures, argues that seven amino acids, four amino terminal and three carboxyl terminal to the scissile bond, are the local determinants of cleavage sites.

Proteases of retrotransposons are less extensively characterized than those of retroviruses. The protease predicted from the gypsy-like element 17.6 protein sequence contains Asp-Thr-Gly at the position of the predicted PR active site, but other retrotransposons, including the intracisternal type A particle, *copia*, and the *copia*-like Ty1 and Ty2 and gypsy-like Ty3 elements of *S. cerevisiae* (3), have Asp-Ser-Gly at the protease active site. The *copia* element encodes a protease which is expressed together with structural proteins from a spliced mRNA. A missense mutation in the active site of this protease prevents accumulation of viruslike particle (VLP) protein, presumably because of instability of the

* Corresponding author.

mutant precursors *in vivo* (60). The yeast Ty1 element encodes an aspartyl protease of 23 kDa at the beginning of the second open reading frame (ORF). It is expressed as part of a polyprotein generated by translational read-through from the first ORF (14, 61). A linker insertion mutation near the putative active site in Ty1 protease blocks processing of the precursor polyprotein to mature Ty1 proteins and results in accumulation of immature particles, although a novel protein is produced from the precursor via either residual Ty1 protease activity or the action of a cellular protease (14, 61). Because processing sites have not been characterized for any retrotransposon protease, the similarity of cleavage specificity between retrotransposon and retroviral proteases is not known.

Ty3 occurs in one to five copies in typical haploid laboratory strains and is composed of a 4.7-kb internal domain flanked by 340-bp long terminal repeats. Ty3 is transcribed into a 5.2-kb genomic RNA (8). The RNA contains two ORFs, *GAG3* of 307 codons, predicted to encode a protein of 290 residues, and *POL3* of 1,270 codons (17). In cells expressing high levels of the 5.2-kb RNA, Ty3 proteins are synthesized and intracellular particles of 156S are observed (15). These particles are primarily composed of structural proteins, encoded by *GAG3*, of 39, 38, 31, 26, and 9 kDa. The 26- and 9-kDa proteins have been designated capsid (CA) and nucleocapsid (NC), respectively. Ty3 particles also contain catalytic proteins encoded by *POL3*, including PR, reverse transcriptase (RT), and integrase (IN), which are homologous to their retroviral counterparts. RT has an apparent mass of 55 kDa, and the IN species have apparent masses of 61 and 58 kDa. *GAG3* proteins are synthesized in relative abundance, while *POL3* proteins, produced from a *GAG3-POL3* fusion polyprotein, are present in the particle at about 1/13 the level of the structural proteins. These particles copurify with Ty3 RNA, as well as RT activity and full-length Ty3 DNA. These particles are analogous to the retroviral core and are referred to as VLPs.

This study was undertaken to (i) determine the effect on particle morphogenesis of inactivating the Ty3 PR, (ii) determine the functional significance of heterology at the active sites of retroviral and retrotransposon PR species, and (iii) identify Ty3 PR cleavage sites in the *GAG3* and *POL3* polyprotein domains.

MATERIALS AND METHODS

Yeast and bacterial strains and culture conditions. Culturing and transformation of *E. coli* and *S. cerevisiae* strains was according to standard methodology (1). *S. cerevisiae* TMy16 (*MATa trp1-H3 ura3-52 his3-Δ200 ade2-101 lys2-1 leu1-12 can1-100 ΔTy3 GAL3⁺*) (39), a derivative of yVB110 which contains no endogenous Ty3 elements (16), was used for experiments in which Ty3 transposition was monitored. Immunoblot analysis was performed on strain AGY-9 (*MATa ura3-52 his4-539 lys2-801 trp1-Δ63 leu2-Δ1 spt3*) (a gift from A. Gabriel and J. Boeke, The Johns Hopkins University). AGY-9 transformed with plasmids expressing Ty3 was used for VLP protein analysis because the mutation in *SPT3* reduces Ty1 expression. AGY-9 also produces higher relative levels of Ty3 proteins than the Ty3 null strain yVB110, from which TMy16 was derived. Processing of Ty3 proteins is indistinguishable in the two strains. *E. coli* RZ1032 (*lysA[61-62] thi-1 relA1 spoT1 dut-1 ung-1 [Tet^r] supE44*) was used for production of single-stranded DNA for site-directed mutagenesis by the method of Kunkel (29). Plasmids were amplified in HB101 (*F⁻ hsd-20 [r_B⁻ m_B⁻]*

recA13 leuB6 ara-14 proA2 lacY1 galK2 rpsL20 [Sm^r] xyl-5 mtl-1 supE44 λ⁻).

VLP preparation. In order to study Ty3 proteins, yeast strain AGY-9 was transformed with pEGTy3-1, a plasmid carrying wild-type Ty3 fused to the *GAL1-10* upstream activating sequence (16). This plasmid also carried the yeast *URA3* gene, sequences from the yeast 2 μ m plasmid which confer high copy number, and the bacterial Amp^r gene and *ori*. Transformed cells were pregrown in medium containing raffinose (a nonrepressing carbon source) and lacking uracil. Cells were transferred to synthetic medium containing galactose as the carbon source and lacking uracil to induce transposition. VLPs were partially purified by using a method developed from that used by Eichinger and Boeke (13) and previously described (15). Briefly, cells were grown to an optical density at 600 nm of 0.9 to 1.1, washed in buffer, digested with zymolyase, lysed with glass beads, and fractionated over a 26-ml 70%–30%–20% sucrose step gradient of 5, 5, and 16 ml, respectively, by centrifugation in a SW28 rotor at 83,000 \times g for 3 h at 4°C. Four milliliters from the 70%–30% interface of each gradient was pooled, concentrated by centrifugation in a Ti50 rotor at 100,000 \times g for 1 h at 4°C, and resuspended in 100 μ l of buffer B (containing Mg²⁺) (13) with 10% glycerol. Protein concentrations were determined by the Bradford assay (4), and purity was assessed by fractionation of 10 μ g of each VLP preparation by electrophoresis on a sodium dodecyl sulfate (SDS)–10% polyacrylamide gel and staining with Coomassie blue.

Immunoblot analysis. Proteins fractionated on polyacrylamide gels were transferred electrophoretically to Nitro-Screen West membranes (DuPont, New England Nuclear Corp.) and analyzed as described previously (15, 17). *GAG3* proteins were visualized with an antibody against a peptide representing residues 30 to 44 of the predicted *GAG3* protein and lying within CA. *POL3* proteins were visualized with an antibody against a peptide encoded by *POL3* codons 1086 to 1099 and lying within IN. The use of these antibodies has been described previously (15, 17).

Amino-terminal sequence analysis of Ty3 proteins. Conditions and techniques used for preparing protein for sequencing were as recommended by P. Matsudaira (36). VLP protein isolated from approximately 30 liters of galactose-induced cells, as described above, was concentrated by centrifugation in a Ti50 rotor at 100,000 \times g for 2 h at 4°C and resuspended in 250 μ l of buffer B with 10% glycerol. A one-third volume of 4 \times Laemmli sample buffer (31) was added. The sample was heated at 37°C for 10 min and centrifuged at 12,000 \times g for 20 min at 4°C to remove insoluble material. An SDS–10% polyacrylamide gel (1 mm by 15 cm by 10 cm) with four 1-cm wells was made by using deionized acrylamide and allowed to polymerize overnight. Four equal aliquots of 33 μ l (about 330 μ g) each were loaded. The running buffer was supplemented to 0.01% thioglycolate. The Ty3 proteins were fractionated by electrophoresis at 200 V for approximately 4.5 h and transferred to a polyvinylidene difluoride (PVDF) membrane (Bio-Rad) for 2.5 h at 400 mA in 10 mM CAPS [(3-(cyclohexylamino)-1-propanesulfonic acid; pH 11.0]–0.1% EDTA–10% methanol. The membrane was then rinsed twice with double-distilled water, stained for 1 min in 0.1% Coomassie blue and 40% methanol, destained in three changes of 50% methanol, and rinsed with water. The bands on the PVDF membrane containing the 61 (IN)-, 58 (IN)-, 55 (RT)-, 38 (Pr38^{*GAG3*})-, and 31 (p31^{*GAG3*})-kDa Ty3 proteins were excised, and the proteins were subjected to five or seven cycles of Edman degradation performed on the Applied Biosystems model

470-A pulsed-liquid protein sequencer by G. Hathaway at the Biotechnology Instrumentation Facility at the University of California, Riverside. A similar procedure employing an SDS-15% polyacrylamide gel was used to fractionate the lower-molecular-mass Ty3 proteins, the 26 (CA)-, 9 (NC)-, and 16 (PR)-kDa species, from a second large-scale preparation of pooled Ty3 VLP protein representing 6 liters of galactose-induced cells. Electrophoresis was for 5 h at 200 V. Electrophoretic transfer was for 2 h at 100 V, but was otherwise performed as described above. The PVDF membrane to which these proteins were transferred was also stained and destained, and the bands containing Ty3 proteins were excised. The proteins were subjected to five cycles of Edman degradation sequencing as described above.

Recombinant DNA manipulations. A pIBI20 plasmid containing the 5' half of Ty3 on a 3.2-kbp *HindIII-SalI* fragment (2) was used for mutagenesis. Two 19-mer oligonucleotides were used to separately mutagenize the sequences that encode the conserved Asp and Ser of the PR active site. Oligonucleotide 212 (5'GATCCACTGATAAACAGGG3') was used to change two nucleotides (underlined), converting the encoded Asp residue to an Ile residue (Asp-59 → Ile). Oligonucleotide 211 (5'GGGTGATCCAGTGTCAAAC3') was used to change a single nucleotide, converting the encoded Ser residue to a Thr residue (Ser-60 → Thr). Plasmid DNA was isolated from transformants and screened by sequence analysis. The 2.0-kbp *BglII-SalI* fragment of Ty3 containing the mutated sequence was then substituted into pEGTy3-1 for the wild-type fragment in the 5' portion of Ty3. These plasmids are referred to as pEGTy3-1mut212 (Asp-59 → Ile) and pEGTy3-1mut211 (Ser-60 → Thr) and were used for protein analysis in AGY-9. The 2.0-kbp *BglII-SalI* fragment was also cloned into pJK312AC, replacing the wild-type fragment in the 5' portion of the Ty3 carried on this plasmid. These plasmids, carrying mutant Ty3 elements, are referred to as pJK312ACmut211 and pJK312ACmut212. The plasmid pJK312AC is identical to pJK311AC (22) in having the yeast *TRP1* gene and *ARS1* and *CEN4* sequences for selection and maintenance at low copy number and the *E. coli* Amp^r gene and *ori*. The Ty3 element on pJK312AC, however, has a deletion of the primer binding site, which prevents replication. This Ty3 is also fused to the *GAL1-10* upstream activating sequence, expresses all Ty3 proteins, and (like pJK311AC) can complement transposition of a marked donor Ty3 (see Results).

RT assays. RT activity was measured under conditions optimized for Ty3 RT (15). VLPs were prepared as described above from galactose-induced AGY-9 cells containing pEGTy3-1, pEGTy3-1mut211, or pEGTy3-1mut212 or no plasmid. RT activity was calculated as the percent incorporation of [α -³²P]dGTP at 26°C in 2.0 h by using a poly(C) template and an oligo(dG) primer. Incorporation of radioactivity was linear during this time for proteins in the range of 1 to 4 μ g in a 30- μ l reaction volume (7, 15).

Transposition assays. Transposition assays were performed by using a helper Ty3 (pJK312AC, or a derivative, pJK312ACmut211 or pJK312ACmut212), incapable of transposition but able to provide proteins in *trans* and a donor Ty3 (pJK314 or pJK422) (22), marked with the yeast *HIS3* gene and incapable of autonomous transposition but containing the *cis*-acting sequences required for transposition. The plasmid pJK314 carried a donor Ty3 in which NC, PR, RT, and IN coding sequences were replaced with *HIS3*, and pJK422 carried a donor Ty3 in which only IN coding sequences were replaced with *HIS3*. Each donor plasmid contained a 2 μ m sequence which confers high copy number

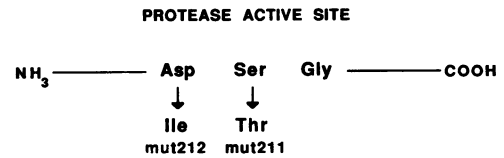


FIG. 1. Mutagenesis of conserved positions in the Ty3 PR active site. The Asp-to-Ile mutation in Ty3-1mut212 and the Ser-to-Thr mutation in Ty3-1mut211 are shown.

in *S. cerevisiae*, the yeast *URA3* gene for selection, and the pIBI20 vector for replication and selection in *E. coli*. Yeast strain TMy16 was transformed separately with each of the three helper plasmids. These transformants were then transformed with one of the donor plasmids. TMy16 was also transformed with each donor plasmid alone. Two individual transformants for each donor or donor-helper combination were grown for 5 days on synthetic media lacking uracil or both uracil and tryptophan, respectively, to select for the retention of these plasmids. The media also contained an appropriate carbon source to induce Ty3 transcription (galactose) or to repress Ty3 transcription (glucose). Five purified colonies representing an original transformant were then patched onto complete medium (YPD) and grown for 1 day to allow loss of the marked donor plasmid. Patches were then replica plated onto synthetic medium lacking histidine and containing 5-fluoro-orotic acid (5FOA) to select for cells which had lost the *URA3*-marked donor plasmid but retained the *HIS3*-marked Ty3.

RESULTS

An Asp-to-Ile mutation in the active site of Ty3 PR blocks Ty3 processing. The similarity among the *copA*, Ty1, and Ty3 retrotransposons and retroviruses and their respective proteases argued that Ty3 PR would be found to be required for maturation of the Ty3 VLP. If immature Ty3 particles accumulated in cells expressing a Ty3 with a mutation in the active site of PR, it would not only implicate PR in the maturation process but would facilitate studies of precursor proteins. Site-directed oligonucleotide mutagenesis was used to change the Asp within the active site consensus Asp-Ser-Gly to Ile (Fig. 1), creating Ty3-1mut212.

To obtain high-level expression of the Ty3 PR mutant, Ty3-1mut212, it was cloned on a high-copy-number plasmid under regulation of the *GAL1-10* upstream activating sequence, as described for pEGTy3-1. Strain AGY-9 was transformed with pEGTy3-1 and, separately, with pEGTy3-1mut212. Transformants were grown in medium lacking uracil and containing galactose, to select for growth of cells retaining the plasmid and to induce expression of the Ty3 elements. Extracts of cells expressing wild-type and mutant Ty3 elements were then fractionated over 70%–30%–20% sucrose step gradients, and the particulate material in the 70%–30% interface fraction was collected by centrifugation. Proteins were fractionated on SDS-10% polyacrylamide gels and identified by Coomassie blue staining (Fig. 2A) and by immunoblot analysis with antibodies against determinants within the CA and IN domains (Fig. 2B and C). The antibodies raised against the CA determinant reacted with a minor p39^{GAG3-POL3} component and with Pr38^{GAG3}, p31^{GAG3}, and CA encoded by wild-type Ty3-1. These proteins were previously identified as GAG3 proteins produced by Ty3-1 (15). The Ty3-1mut212 showed Pr38^{GAG3} at low levels and an even lower amount of a protein with an

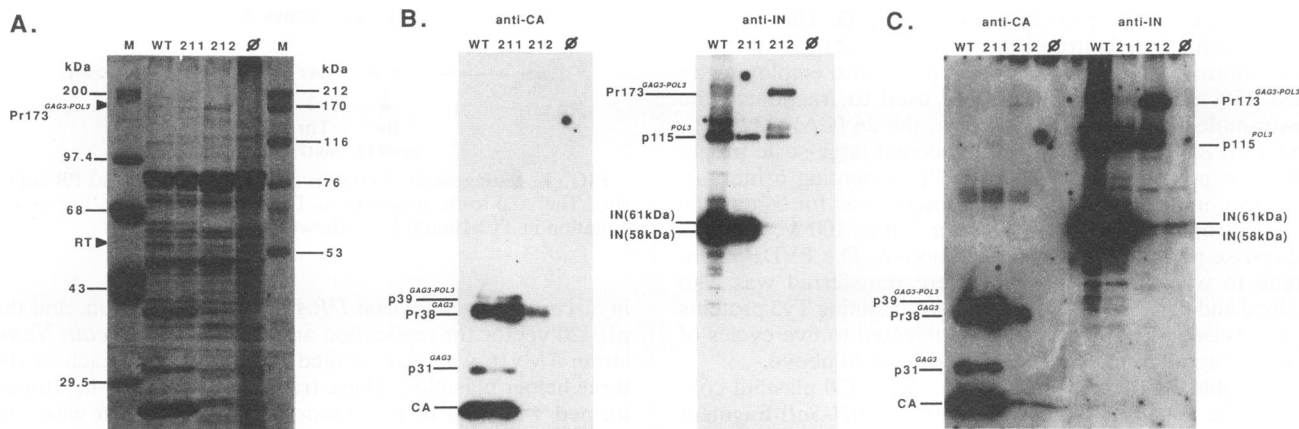


FIG. 2. Analysis of VLP proteins isolated from cells expressing Ty3 PR mutants. (A) Coomassie blue-stained SDS-10% polyacrylamide gel in which a 40- μ g aliquot of the particulate fraction was loaded per lane. AGY-9 cells transformed with pEGTy3-1 (WT), pEGTy3-1mut211 (211), and pEGTy3-1mut212 (212) or with AGY-9 cells alone (\emptyset) were sources of the particulate fractions. Molecular mass standards used to determine the size of the *GAG3-POL3* polyprotein are shown in flanking lanes (M), and the molecular masses are indicated on both sides of the gel. Ty3 proteins Pr173^{GAG3-POL3} and RT are labeled. (B) Immunoblot analysis of the 70%-30% interface fraction. Extracts are from cells defined above for panel A. Ten micrograms of protein was fractionated on an SDS-10% polyacrylamide gel (duplicate samples were loaded in two lanes), transferred to nitrocellulose, probed with either anti-CA or anti-IN antibodies, reacted with ¹²⁵I-protein A, washed, and exposed to X-ray film for 6 days. The Ty3 proteins are labeled. (C) Four-week exposure of the immunoblots used for exposures shown in panel B.

apparent mass of 173 kDa (Fig. 2), similar to the predicted size of a *GAG3-POL3* fusion protein (175 kDa). The p39^{GAG3-POL3}, p31^{GAG3}, and CA proteins were absent. CA and the 61- and 58-kDa IN species were detected in Ty3-1mut212 only after overexposure of the immunoblot (Fig. 2C) and are present at the same level as in the control sample (nontransformed cells). Strain AGY-9 has four endogenous copies of Ty3 (39), and expression of these elements is likely to account for this background.

Antisera raised against the IN determinant reacted with a 115-kDa species and 61- and 58-kDa IN species in the extracts from the cells containing the wild-type Ty3-1 (17) (Fig. 2B and C). In extracts of cells expressing the Ile-Ser-Gly active-site mutant, the most prominent reactive protein had an apparent mass of 173 kDa (Fig. 2B and C). We were also able to detect a species of approximately 115 kDa which may have been generated by residual PR activity, the action of a cellular protease, or the action of a contaminating protease(s) introduced during VLP preparation. There were much lower overall levels of proteins which reacted with the antibody against IN in extracts of cells expressing the mutant compared with those in extracts of cells expressing the wild-type Ty3. Thus, inactivation of Ty3 PR does not result in accumulation of immature Ty3 particles to the levels observed for mature particles in cells expressing wild-type Ty3 elements. This could be due either to less efficient formation of the mutant particles or more rapid turnover of the particles that form.

PR is required for wild-type levels of RT activity and for transposition. The particulate fractions from cells expressing wild-type and mutant Ty3 elements were assayed for RT activity (Fig. 3). The level of [α -³²P]dGTP incorporation mediated by extracts of cells expressing pEGTy3-1mut212 was slightly above background and approximately 10% that of extracts expressing wild-type Ty3 in the linear range of 1 to 4 μ g of protein. Because no mature RT was detected by Coomassie blue staining in extracts of this PR mutant (Fig. 2A), we speculate that the source of the activity over background levels observed in nontransformed cells is the

115-kDa protein which reacts with the anti-IN antisera. The size of this protein is consistent with that of an RT-IN fusion. Alternatively, it is possible that RT activity is reduced severalfold, rather than being completely absent, in immature particles composed of unprocessed polyproteins, similar to what has been shown for a human immunodeficiency virus PR active-site mutant (47) and for Ty1 PR active-site mutants (61).

The absence of mature proteins in the PR mutant particles and the reduced level of Ty3 proteins in general suggested that the mutation would block Ty3 transposition, despite

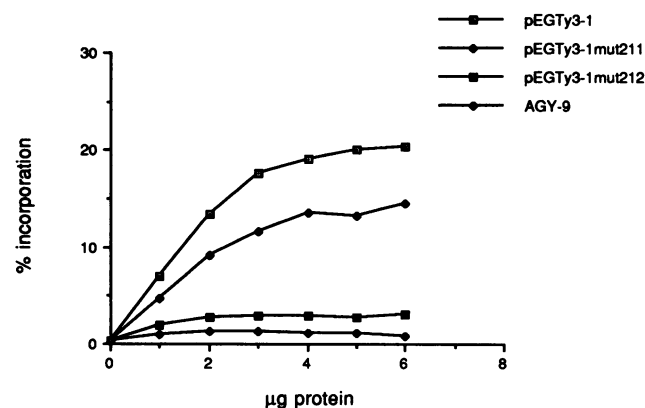


FIG. 3. RT assays of extracts from cells expressing wild-type and PR mutant Ty3 elements. AGY-9 cells transformed with high-copy-number Ty3 plasmids or not transformed were grown in galactose-containing, uracil-lacking media, and extracts of these cultures were fractionated over sucrose step gradients. One to 6 μ g of total VLP protein was used in each assay for each sample, as described in Materials and Methods. RT activity is expressed as the percent incorporation of [α -³²P]dGTP. These values are the averages of duplicate assays and are representative of assays performed with multiple samples.

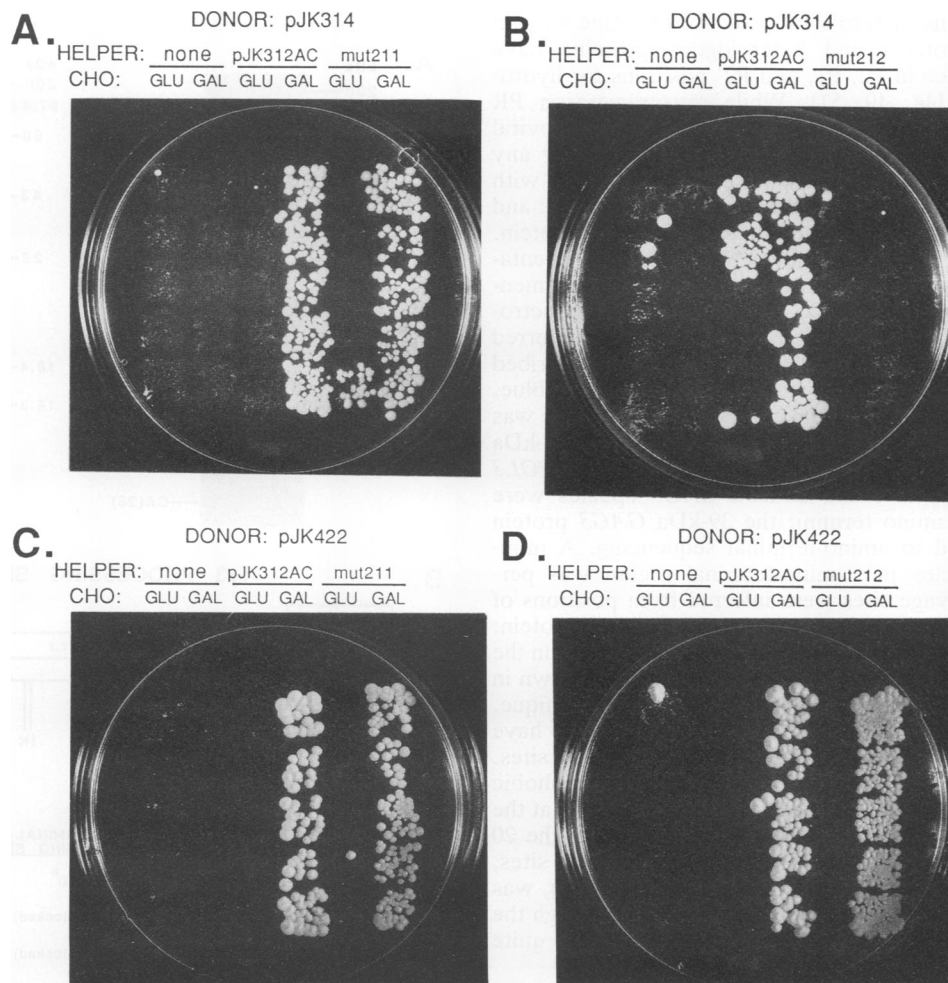


FIG. 4. Transposition assays of Ty3 PR mutants. TMy16 cells were transformed with pJK312AC, pJK312ACmut211, and pJK312ACmut212 (helper Ty3 plasmids) and then with pJK314 (donor Ty3 plasmid not encoding PR) or with pJK422 (donor Ty3 plasmid encoding PR). TMy16 cells were also transformed with each donor plasmid alone. Transformants were streaked onto selective synthetic media containing either glucose (GLU) (to repress transposition) or galactose (GAL) (to induce transposition). Five isolated colonies of each were then patched onto YPD nonselective media to allow loss of the donor plasmid and replica plated onto media containing 5FOA and lacking histidine to select for the acquisition of a genomic marked donor Ty3. Growth on medium lacking histidine and containing 5FOA is shown. The plasmids contained in the transformant and the carbon source on which cells were grown are indicated above each column of patches on the plates.

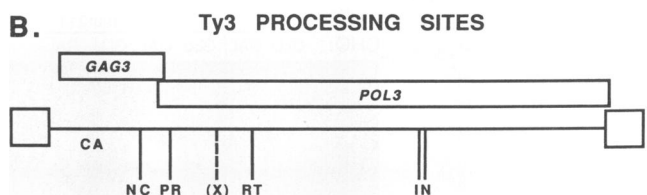
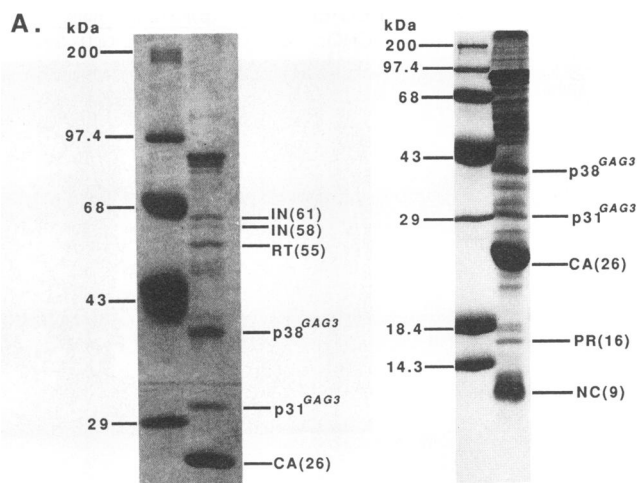
detectable RT activity. This prediction was tested by transforming the Ty3 null strain TMy16 with a galactose-inducible, wild-type, or mutant helper Ty3 element, carried on a low-copy-number plasmid, pJK312AC or pJK312ACmut212, respectively, and with a galactose-inducible donor Ty3 element carrying a *HIS3* insertion on a high-copy-number plasmid, pJK314 or pJK422. A wild-type helper Ty3 supplies proteins required for transposition of a donor Ty3 element, thereby complementing its transposition (6, 22) (see Materials and Methods). In this assay, cells carrying the plasmids are induced for transposition on galactose-containing medium lacking uracil and tryptophan and are patched onto YPD to allow plasmid loss. Patches of cells are transferred onto medium containing 5FOA to select against cells containing the *URA3*-marked donor plasmid and lacking histidine to select for cells which have acquired a genomic copy of the *HIS3*-marked donor Ty3. Testing of the Ile-Ser-Gly PR mutant helper Ty3 in this assay with a donor lacking the PR coding region (pJK314) showed that the PR mutant did

not complement transposition of this donor Ty3 (Fig. 4B). The level of growth on medium lacking histidine and containing 5FOA was the same, whether regrowth was on an inducing or noninducing carbon source. It was possible that transposition was blocked because the mutation caused the Ty3 polyprotein to adopt an aberrant structure which was a poor substrate for processing. To assess this possibility, the donor Ty3 element in plasmid pJK422, which encoded PR but lacked most of the IN coding sequence, was used to test whether Ty3-1mut212 could be rescued by wild-type PR. Transposition of this donor did occur (Fig. 4D). Thus, the *GAG3-POL3* polyprotein containing the mutant PR could be processed to produce at least the IN species required for transposition. This suggested that the absence of transposition observed for pJK314 together with pJK312ACmut212 was due to lack of PR activity, rather than inability of the polyprotein to be incorporated into particles or to be a substrate for processing.

Ty3 polyprotein cleavage sites are within hydrophobic re-

gions. Comparisons of retroviral protease cleavage sites in retroviral polyproteins and heterologous proteins have shown a preference in P2, P1, and P2' positions for hydrophobic residues (44, 49, 51). While retrotransposon PR species share two conserved regions with those of retroviral PR species, cleavage sites have not been defined for any retrotransposon protease. AGY-9 cells transformed with pEGTy3-1 were grown in medium containing galactose and lacking uracil to obtain large amounts of Ty3 VLP protein. The cell extracts were fractionated by velocity sedimentation over a sucrose step gradient, concentrated by sedimentation, and fractionated by SDS-polyacrylamide gel electrophoresis (PAGE) (Fig. 5A). The proteins were transferred from polyacrylamide gels to a PVDF membrane as described in Materials and Methods and stained with Coomassie blue. Bands were excised, and the amino-terminal sequence was determined by automated Edman degradation for the 9-kDa GAG3 protein and the 16-, 55-, 61-, and 58-kDa POL3 proteins. The 38-, 31-, and 26-kDa GAG3 species were blocked at their amino termini; the 39-kDa GAG3 protein was not subjected to amino-terminal sequencing. A minimum of five cycles per amino-terminal region was performed. The cleavage sites were inferred from positions of amino-terminal sequence in the predicted Ty3 polyprotein; each amino-terminal sequence occurred only once within the predicted Ty3 protein sequence. These results are shown in Fig. 5B. Although each putative processing site is unique, two sites have His residues at the P1 position and two have Ser residues. Thr is the P1' residue in three of the five sites. The P2 and P2' positions are occupied by hydrophobic residues: Val at four of the five P2 sites, and Val or Ile at the five P2' sites. The hydrophobicity index of each of the 20 amino acids surrounding each of the five cleavage sites, determined by the method of Kyte and Doolittle (30), was summed and plotted (Fig. 6). It is apparent that although the P1 and P1' positions are not, P3, P2, and P2' are quite hydrophobic.

Changing the PR active-site Ser to Thr does not result in significant changes in the levels or sizes of Ty3 proteins or in the ability of Ty3 to transpose. The conserved active site of most retroviral PR proteins is Asp-Thr-Gly. Although avian sarcoma/leukosis virus which has Asp-Ser-Gly at the active site, has been shown to be capable of cleaving heterologous substrate proteins and peptides (12, 21, 25, 48), whether Ser and Thr are fully interchangeable against the homologous substrate *in vivo* has not been tested. The Ser in the Ty3 PR active-site Asp-Ser-Gly sequence was changed to Thr to produce the mutant element Ty3-1mut211 (Fig. 1). Extracts of cells expressing this mutant were fractionated on 70%–30%–20% sucrose step gradients and further separated on SDS–10% polyacrylamide gels, and then stained with Coomassie blue (Fig. 2A) or transferred to NitroScreen West for immunoblot analysis with antisera against CA and IN domains (Fig. 2B and C). GAG3 proteins were present at similar levels compared with their counterparts in the wild-type Ty3. However, Western blot (immunoblot) analysis showed that p115 and the 61- and 58-kDa IN species were slightly underrepresented in particles from mutant compared with wild-type Ty3 elements. RT assays showed activity in the extracts of cells expressing Ty3-1mut211 to be approximately 61% that of wild-type Ty3 within the linear range of 1 to 4 μ g of protein (Fig. 3). The Coomassie blue-stained proteins included a species of the correct size for RT (55 kDa) present at reduced levels in the mutant Ty3 sample (Fig. 2A). The simplest interpretation of the results of the RT assay is, therefore, that the decrease in activity is due to the



PROTEIN	MOL. MASS (kDa)		N-TERMINAL PROCESSING SITE		P1/P1' RESIDUES
	PRED.	APP.			
p39 ^{GAG3-POL3}	38	39	ND ^b		ND
p38 ^{GAG3}	34	38	ND (blocked)		ND
p31 ^{GAG3}	NA ^a	31	ND (blocked)		ND
CA	NA	26	ND (blocked)		ND
			P5 P4 P3 P2 P1 ↓ P1' P2' P3' P4' P5'		
NC	7	9	N	G Y V H T V R T R	233/234 ^c
PR	25	16	L	P I V H Y I A I P	32/33 ^d
RT	55	55	S	N V V S T I Q S V	258/259 ^d
IN	62	61	S	R A V Y T I T P E	734/735 ^d
	59	58	D	P L C S A V L I H	760/761 ^d

^a Not applicable

^b Not determined

^c Residues within the predicted GAG3 polyprotein

^d Residues within the predicted POL3 polyprotein

FIG. 5. Ty3 proteolytic processing sites. (A) SDS-PAGE analysis of aliquots of Ty3 VLP protein used for Edman degradation sequencing. Ten to 15 μ g of Ty3 VLP protein produced in cells transformed with pEGTy3-1, induced by growth in galactose-containing medium, was loaded per lane on either a 10% polyacrylamide (left)– or a 15% polyacrylamide (right)–SDS gel. *M_r* standards are indicated on the left, and Ty3 proteins are labeled on the right. (B) Schematic of Ty3 with putative N-terminal processing sites shown for NC, PR, RT, and IN (61- and 58-kDa species), a comparison of predicted (PRED.) and apparent (APP.) molecular masses for all Ty3 proteins, and the five putative cleavage sites (P5 to P5') deduced from the amino-terminal sequence. Predicted value is the mass calculated assuming no additional processing sites in the GAG3-POL3 polyprotein; (X) indicates the putative N-terminal processing site of protein X (see Discussion). Apparent molecular masses were derived by comparison to the mobility of standards.

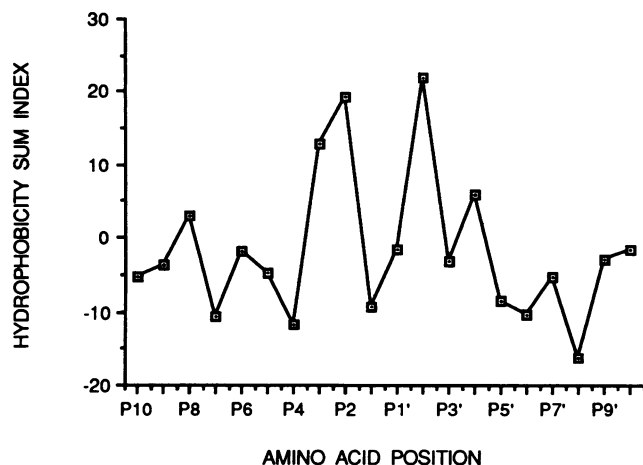


FIG. 6. Plot of hydrophobicity indices of Ty3 PR cleavage contexts. The hydrophobicity was determined by the method of Kyte and Doolittle (30) with a DNA sequence analysis program written by A. Goldin and G. Gutman (University of California, Irvine) from position P10 to position P10' for the five Ty3 PR putative sites of cleavage and summed at each position. A similar analysis was previously presented for retroviral PR cleavage sites (49).

reduced level of RT in this fraction, rather than to reduced specific activity of the mature RT.

The ability of a helper Ty3 with the Ser-to-Thr mutation to complement transposition was also assayed. By this qualitative assay, the level of transposition of a marked donor Ty3, either encoding PR or not, was the same when complemented by either the wild-type helper Ty3 element or the Asp-Thr-Gly PR mutant helper Ty3 (Fig. 4A and C). Thus, in contrast to results with the Ile-Ser-Gly active-site mutant, function of the Asp-Thr-Gly mutant PR was not limiting for transposition.

DISCUSSION

The Ty3 PR active site was identified by comparison of the sequence of the predicted Ty3 *POL3* protein to sequences of cellular aspartyl proteases and proteases of retroviruses and retrotransposons (16). The mutation changing Asp to Ile in the active site of PR eliminated or reduced the levels of mature Ty3 proteins, implicating PR in processing of Ty3 *GAG3* and *GAG3-POL3* polyproteins. This phenotype is inferred to result directly from disruption of PR function, rather than indirectly from conversion of the *GAG3-POL3* polyprotein to a form which assumes an aberrant conformation. When the Ty3 having the Asp-to-Ile mutation in PR was expressed as a helper in *trans* with a donor Ty3 which encoded the wild-type PR, processing and transposition occurred. Thus, the conformation of the mutant Gag3-Pol3 was compatible with sufficient processing for transposition to occur. The variety of mutations in retrotransposon (60, 61) and retroviral PR active sites which confer similar phenotypes also argues that the block in maturation observed is unlikely to result from specific structural effects (9, 10, 20, 23, 46, 47, 55, 56). In our study, levels of RT activity in the mutant were slightly above the background observed in nontransformed cells. Whether this resulted from a low level of activity of the Gag3-Pol3 precursor or generation of some active species by background levels of processing, either by uninduced, endogenous Ty3-encoded PR or by

cellular proteases, cannot be determined from these data. The 115-kDa species which reacts with the anti-IN antibody is present in the PR mutant and almost certainly contains the RT domain. Differing amounts of RT activity have been observed for PR mutants in other systems. It seems likely that these differences reflect both the particular mutations used to inactivate PR and the system. In the case of Ty1, three mutations which eliminated detectable processing and RT activity measured by an endogenous *in vitro* assay did not interfere with activity measured by an exogenous *in vitro* assay (61). One of the mutants which contained a large deletion of the PR and IN coding sequences actually enhanced this activity about 15-fold. In avian retroviruses, mutagenesis of the PR active site virtually eliminated RT activity (9, 55, 56). In murine leukemia virus (10, 20) and human immunodeficiency virus (46, 47), significant *in vitro* activity has been reported for some, but not all, PR mutants.

Although existence of a Ty3 Gag3-Pol3 fusion protein was previously inferred from experiments in which *POL3* was fused in frame to *lacZ* (22), the experiments reported here allowed direct visualization and identification of the predicted *GAG3-POL3* polyprotein for the first time. The Ty3 Gag3-Pol3 appears to be a labile species. In cells expressing either a *GAG3-POL3* fusion mutant (22) or a PR active-site mutant Ty3 under control of a strong promoter, the *GAG3-POL3* polyprotein did not accumulate in immature particles. Analysis of extracts from cells expressing the *GAG3-POL3* fusion mutant Ty3 in linear sucrose gradients failed to show precursor species in any fraction, indicating that these proteins are also not accumulating as nonparticulate species. These results contrast with the results of experiments which showed accumulation of polyprotein precursors and formation of immature particles when proteolytic activity was disrupted by a variety of mutations in the PR active site in retroviruses or Ty1 retrotransposons (10, 20, 33, 47, 55, 61). This result was similar, however, to that seen with *copA* structural-catalytic protein fusion and PR active-site mutants which did not accumulate unprocessed polyproteins (60). There may be differences in the accessibility of different polyproteins to cellular proteases. It does not seem likely that the failure of precursor proteins and immature particles to accumulate is due to disruption of the normal structure in the case of either the Ty3 PR mutant or the *GAG3-POL3* fusion mutant, because both can be complemented by donors which are not themselves transpositionally competent (22; this study). A model consistent with these findings is that the *GAG3-POL3* polyprotein is an unstable, transient species, possibly even under conditions in which is incorporated into particles, because of either specific processing or nonspecific degradation.

In general, the masses predicted for proteins generated by cleavage of *GAG3* and *GAG3-POL3* polyproteins at the positions identified here are consistent with the apparent masses of these proteins (Fig. 5B). The protein predicted for translation of *GAG3* from the putative initiation Met through the end of the ORF is 290 residues in length. A major Pr38^{*GAG3*} and a minor p39^{*GAG3-POL3*} species are candidates for the full-length *GAG3* protein. Mutation of the protease resulted in loss of the 39-kDa species, as well as the 26-, 31-, and 9-kDa species, but not the 38-kDa species. This suggests that the p39^{*GAG3-POL3*} species is processed from the *GAG3-POL3* fusion protein—possibly at the amino terminus of PR. The observation that the synthetic Ty3 *GAG3-POL3* fusion mutant produced no Pr38^{*GAG3*} but did produce p39^{*GAG3-POL3*} protein (22) is consistent with this interpretation.

The amino termini of the 38-, 31-, and 26-kDa proteins

were blocked. These proteins all reacted with the anti-CA antibody against a peptide representing residues 30 to 44 of the predicted *GAG3* protein and failed to react with an antibody raised against a peptide representing *GAG3* codons 235 to 252 (15). The amino-terminal location and the sizes of these three proteins are consistent with common amino termini. The 31-kDa protein could be generated by the cleavage between residues 232 and 233, which defines the amino terminus of NC, and the 26-kDa species could be produced by additional processing. Alternatively, the carboxyl terminus of the 26-kDa protein could be determined by the NC amino-terminal cleavage and the 31-kDa produced by modification of the 26-kDa species. Inspection of the protein sequence in the region which might be cleaved to generate the carboxyl terminus of a 26-kDa species from a 31-kDa species terminating at NC has allowed us to identify the sequence Thr-Thr-Ala-Leu-Thr-Glu-Arg (*GAG3* codons 185 to 191) (where Leu and Thr could occupy the P1 and P1' positions, respectively) which would be compatible with the features common to known retroviral and Ty3 processing sites. Assuming all three *GAG3* proteins have a common amino terminus, the most likely explanation of a blocked amino terminus is acetylation. The amino-terminal sequence of the predicted *GAG3* protein is Met-Ser, a sequence which is subject to processing and N-acetylation in *S. cerevisiae* (18, 32). This modification of a structural polyprotein is not unprecedented, as the amino terminus of Pr76^{gag} of avian sarcoma/leukosis virus was shown to be acetylated (43).

The NC protein would have a maximum of 57 residues if the carboxyl terminus is defined by the end of *GAG3*. Although this is fewer than would be expected on the basis of the apparent mass, retroviral NC proteins, which are close to 60 amino acids in length and similar in composition to Ty3 NC, migrate anomalously slowly, perhaps because of their highly basic nature.

The amino-terminal sequence of the 16-kDa protein begins 26 residues from the Asp of the PR active site, similar to the spacing of 24 to 36 residues found among three retroviral PRs (23, 45, 49). A 16-kDa protein reacts weakly with antibodies raised against a peptide from the proposed PR coding region (7). If no other cleavage occurred between the amino terminus of the PR and that of RT, a protein of 228 residues would be generated. Therefore, an additional cleavage site is inferred. Analysis of the protein sequence in the implicated region by using the ANALSEP program (50, 54), operating on a data base consisting of known Ty3 processing sites, identified the sequence Glu-Thr-Val-Asn-Asn-Val-Arg-Thr-Tyr-Ser (*POL3* codons 162 to 171) in which cleavage could occur between either Asn-Asn or Arg-Thr. Cleavage at one of these two positions in the polyprotein would generate a Ty3 PR of 132 or 135 residues whose predicted mass is close to the apparent mass of 16 kDa.

The postulated cleavage that creates the carboxyl terminus of PR would also generate a protein of about 90 residues encoded between the regions which encode PR and RT. This region contains an in-frame direct repeat of 18 bp flanking a 60-bp sequence and encodes a protein sequence with no known similarity to retroviral proteins. In Ty3-2 (16), this 60-bp sequence is repeated, flanked by a total of three copies of the 18-bp sequence. It is known that chimeras of Ty3-1 and Ty3-2 which contain the duplication are transpositionally active (17). Therefore, if this region has a function, it is not disrupted by an additional copy of the insertion. A protein of approximately 10.5 kDa could be generated from this cleavage. We have observed proteins which migrate close to PR in SDS-PAGE analysis of Ty3 VLP prepara-

tions, including one of 11 kDa, but we could not unequivocally distinguish one such protein from the background of cellular proteins present in cells not overexpressing Ty3 and migrating in the same region.

Retroviral proteases are homodimeric proteins, the subunits of which are found as repeated domains in the related monomeric cellular aspartyl proteases. It has been proposed that cellular and retroviral proteases evolved from a common dimeric protein, with eventual generation of a cellular gene encoding the monomeric cellular protease (45). It is tempting to speculate that a protein of approximately 90 amino acids (protein X), encoded downstream of the Ty3 PR, could represent a vestigial version of the cellular aspartyl protease and that, eventually, insertions into this sequence and acquisition of a PR cleavage site completed the transition to a dimeric Ty3 PR. Support for this theory was sought by inspection of the Ty3 PR active-site sequence and the predicted sequence of the postulated protein. The immediate context of the Ty3 PR active site, Phe-Asp-Ser-Gly-Ser, does in fact have closer resemblance to the cellular proteases (Phe/Val-Asp-Thr-Gly-Ser) than to the retroviral proteases (Leu/Val-Asp-Thr-Gly-Ala) (45). Within the downstream coding region, a sequence encoding Asp-Ser occurs 34 or 37 codons from the predicted cleavage site, and the sequence Ile-Leu is encoded 11 codons from the RT amino terminus. This has a weak resemblance to the protease hydrophobic motif. A corollary of the hypothesis that Ty3 PR actually evolved from a monomeric cellular-type aspartyl protease is that there may have been positive selection for viruses which lost one domain of the monomeric PR. Dimerization could then develop as a means of regulating activation. Inspection of both the interstitial type A particle H18 and simian AIDS retrovirus predicted protein sequences has also revealed an in-frame PR-like region adjacent to the PR coding region (protein X) which is flanked by imperfect direct nucleotide repeats (37, 52). Whether these sequences are processed from the polyprotein and have a function is also not known. A modified PR domain could have specific functions, including processing in the absence of a high concentration of retrotransposon proteins, facilitating folding of an active half-domain, or inhibiting premature activity. Whether any of these functions might persist in the current version of Ty3 can be tested by determining the effects of directed mutations within this region on the kinetics of Ty3 PR activity.

The amino terminus of RT is consistent with a 55-kDa protein, the carboxyl terminus of which is defined by the amino terminus of the 61-kDa IN species. The predicted masses for IN proteins with amino termini observed in this study and a common carboxyl terminus defined by the end of *POL3* are 62 and 59 kDa. These are close to the observed values of 61 and 58 kDa. A mutant (Ty3-2) which terminates prematurely in the IN coding region produces species which differ in apparent mass by about the same amount as the mature species (17). This finding, together with the apparent masses of the mature proteins, suggests that the carboxyl termini of the 61- and 58-kDa IN species do not differ from each other and are encoded at the end of *POL3*.

Alignment of retroviral cleavage sites and experiments producing *in vitro* cleavage of synthetic peptides have implicated residues P4 to P3' in specifying PR substrates. A role for a larger context can be inferred from experiments showing that the same cleavage sequence is not an equivalent substrate at different positions in the *gag-pol* polyprotein (34, 58). Comparison of the five Ty3 polyprotein cleavage sites identified here showed no position which was

absolutely conserved. Nevertheless, a hydrophobic residue was consistently present at P3 (Ile, Val, and Leu in three of the five sites), P2 (Val at four out of five sites), and P2' (Ile at three sites and Val at the other two sites). P1 and P1' positions were not found to be occupied by especially hydrophobic residues. This was different from the results of summing the hydrophobicity indices of 46 retroviral sites, which showed that the P3 residue is not particularly hydrophobic but that the P1 and P1' residues are (49). The Ty3 cleavage sites did not encompass all the features from either the type I or type II pattern defined by Pettit et al. (49) in comparing 46 retroviral cleavage sites, nor did any of them precisely match known processing sites of retroviral polyproteins (44, 49). Inspection of the Ty3 *GAG3* and *POL3* predicted protein sequence by using the ANALSEP program and a weight matrix derived from 46 retroviral processing sites, however, did identify Ty3 processing sites for IN 61- and 58-kDa species (50).

In summary, we have identified the Ty3 PR and demonstrated its critical role in VLP maturation and transposition. We have also identified and characterized five of its putative cleavage sites in the Ty3 *GAG3-POL3* polyprotein. From these cleavage sites, we deduced the existence of an additional Ty3 protein(s) encoded between PR and RT coding regions. The cleavage context is, as is the case for retroviruses, relatively hydrophobic. The decreased hydrophobicity that we observed at P1 in Ty3 compared with that observed in retroviruses suggested modest differences between the specificities of Ty3 and retroviral PR. Interestingly, changing the Ser in the Asp-Ser-Gly active-site sequence found in *copiA*, Ty1, Ty3, and Rous sarcoma virus to Thr in Ty3 to mimic the Asp-Thr-Gly sequence found in most retroviruses allowed production of all mature Ty3 proteins. Although the mutant PR appeared to process *POL3* proteins somewhat less efficiently than the wild type, Ser and Thr are both functional in this position in the active site.

ACKNOWLEDGMENTS

We thank G. Hathaway, Biotechnology Instrumentation Facility, for the amino-terminal sequence analysis. We thank D. Loeb for helpful discussions and communicating results prior to publication and R. Swanstrom and S. Pettit for inspection of the *GAG3-POL3* polyprotein sequence for sites predicted to be cleaved by retroviral and Ty3 proteases.

This work was supported by P.H.S. grant GM33281 to S.B.S. J.K. was supported on P.H.S. predoctoral training grant GM07311.

REFERENCES

- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl. 1992. Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, Boston.
- Bilanchone, V. B. Unpublished data.
- Boeke, J. D., and S. B. Sandmeyer. 1991. Yeast transposable elements, p. 193-261. In J. Pringle, E. Jones, and J. Haber (ed.), The molecular and cellular biology of the yeast *Saccharomyces*: genome dynamics, protein synthesis, and energetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248-254.
- Burstein, H., D. Bizub, and A. M. Skalka. 1991. Assembly and processing of avian retroviral *gag* polyproteins containing linked protease dimers. *J. Virol.* **65**:6165-6172.
- Chalker, D. L., and S. B. Sandmeyer. 1990. Transfer RNA genes are genomic targets for *de novo* transposition of the yeast retrotransposon Ty3. *Genetics* **126**:837-850.
- Chalker, D. L., and S. B. Sandmeyer. Unpublished data.
- Clark, D. J., V. W. Bilanchone, L. J. Haywood, S. L. Dildine, and S. B. Sandmeyer. 1988. A yeast sigma composite element, Ty3, has properties of a retrotransposon. *J. Biol. Chem.* **263**:1413-1423.
- Craven, R. C., R. P. Bennet, and J. W. Wills. 1991. Role of the avian retroviral protease in the activation of reverse transcriptase during virion assembly. *J. Virol.* **65**:6205-6217.
- Crawford, S., and S. P. Goff. 1985. A deletion mutation in the 5' part of the *pol* gene of Moloney murine leukemia virus blocks proteolytic processing of the *gag* and *pol* polyproteins. *J. Virol.* **53**:899-907.
- Debouck, C., J. G. Gorniak, J. E. Strickler, T. D. Meek, B. W. Metcalf, and M. Rosenberg. 1987. Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the *gag* precursor. *Proc. Natl. Acad. Sci. USA* **84**:8903-8906.
- Demsey, A., F. Collins, and D. Kawka. 1980. Structure of and alterations to defective murine sarcoma virus particles lacking envelope proteins and core polyprotein cleavage. *J. Virol.* **36**:872-877.
- Eichinger, D. J., and J. D. Boeke. 1988. The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition. *Cell* **54**:955-966.
- Garfinkel, D. J., A. M. Hedge, S. D. Youngren, and T. D. Copeland. 1991. Proteolytic processing of *pol-TYB* proteins from the yeast retrotransposon Ty1. *J. Virol.* **65**:4573-4581.
- Hansen, L. J., D. L. Chalker, K. J. Orliński, and S. B. Sandmeyer. 1992. Ty3 *GAG3* and *POL3* genes encode the components of intracellular particles. *J. Virol.* **66**:1414-1424.
- Hansen, L. J., D. L. Chalker, and S. B. Sandmeyer. 1988. Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. *Mol. Cell. Biol.* **8**:5245-5256.
- Hansen, L. J., and S. B. Sandmeyer. 1990. Characterization of a transpositionally active Ty3 element and identification of the Ty3 integrase protein. *J. Virol.* **64**:2599-2607.
- Huang, S., R. C. Elliott, P.-S. Liu, R. K. Koduri, J. L. Weickmann, J.-H. Lee, L. C. Blair, P. Ghosh-Dastidar, R. A. Bradshaw, K. M. Bryan, B. Einarson, R. L. Kendall, K. H. Kolacz, and K. Saito. 1987. Specificity of cotranslational amino-terminal processing of proteins in yeast. *Biochemistry* **26**:8242-8246.
- Katoh, I., Y. Ikawa, and Y. Yoshinaka. 1989. Retrovirus protease characterized as a dimeric aspartic proteinase. *J. Virol.* **63**:2226-2232.
- Katoh, I., Y. Yoshinaka, A. Rein, M. Shibuya, T. Odaka, and S. Oroszlan. 1985. Murine leukemia virus maturation: protease region required for conversion from "immature" to "mature" core form and for virus infectivity. *Virology* **145**:280-292.
- Khan, A. S., and J. R. Stephenson. 1979. Feline sarcoma virus-coded polyprotein: enzymatic cleavage by a type C virus-coded structural protein. *J. Virol.* **29**:649-656.
- Kirchner, J., S. B. Sandmeyer, and D. F. Forrest. 1992. Transposition of a Ty3 *GAG3-POL3* fusion mutant is limited by availability of capsid protein. *J. Virol.* **66**:6081-6092.
- Kohl, N. E., E. A. Emimi, W. A. Schleif, L. J. Davis, J. C. Heimbach, R. A. F. Dixon, E. M. Scolnick, and I. S. Sigal. 1988. Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. USA* **85**:4686-4690.
- Kotler, M., W. Danho, R. A. Katz, J. Leis, and A. M. Skalka. 1989. Avian retroviral protease and cellular aspartic proteases are distinguished by activities on peptide substrates. *J. Biol. Chem.* **264**:3428-3435.
- Kotler, M., R. A. Katz, W. Danho, J. Leis, and A. M. Skalka. 1988. Synthetic peptides as substrates and inhibitors of a retroviral protease. *Proc. Natl. Acad. Sci. USA* **85**:4185-4189.
- Kotler, M., R. A. Katz, and A. M. Skalka. 1988. Activity of avian retroviral protease expressed in *Escherichia coli*. *J. Virol.* **62**:2696-2700.
- Kramer, R. A., M. D. Schaber, A. M. Skalka, K. Ganguly, F. Wong-Staal, and E. P. Reddy. 1986. HTLV-III *gag* protein is processed in yeast cells by the virus *pol*-protease. *Science* **231**:1580-1584.
- Krausslich, H. G. 1991. Human immunodeficiency virus protein-

- ase dimer as a component of the viral polyprotein prevents particle assembly and viral infectivity. *Proc. Natl. Acad. Sci. USA* **88**:3213-3217.
29. Kunkel, T. A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**:488-492.
 30. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**:105-132.
 31. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**:680-685.
 32. Lee, F.-J., L.-W. Lin, and J. A. Smith. 1990. Model peptides reveal specificity of N-acetyltransferase from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**:11576-11580.
 33. Le Grice, S. F., J. Mills, and J. Mous. 1988. Active site mutagenesis of the AIDS virus protease and its alleviation by trans complementation. *EMBO J.* **7**:2547-2553.
 34. Loeb, D. D., C. A. Hutchison III, M. H. Edgell, W. G. Farmerie, and R. Swanstrom. 1989. Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases. *J. Virol.* **63**:111-121.
 35. Loeb, D. D., R. Swanstrom, L. Everitt, M. Manchester, S. E. Stamper, and C. A. Hutchison III. 1989. Complete mutagenesis of the HIV-1 protease. *Nature (London)* **340**:397-400.
 36. Matsudaira, P. T. 1989. A practical guide to protein and peptide purification for microsequencing. Academic Press, San Diego.
 37. McClure, M. A., M. S. Johnson, and R. F. Doolittle. 1987. Relocation of a protease-like gene segment between two retroviruses. *Proc. Natl. Acad. Sci. USA* **84**:2693-2697.
 38. Meek, T. D., B. D. Dayton, B. W. Metcalf, G. B. Dreyer, J. E. Strickler, J. G. Gorniak, M. Rosenberg, M. L. Moore, V. W. Magaard, and C. Debouck. 1989. Human immunodeficiency virus 1 protease expressed in *Escherichia coli* behaves as a dimeric aspartic protease. *Proc. Natl. Acad. Sci. USA* **86**:1841-1845.
 39. Menees, T., and S. B. Sandmeyer. Unpublished data.
 40. Miller, M., M. Jaskolski, J. K. Rao, J. Leis, and A. Wlodawer. 1989. Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature (London)* **337**:576-579.
 41. Navia, M. A., P. M. D. Fitzgerald, B. M. McKeever, C.-T. Leu, J. C. Heimbach, W. K. Herber, I. S. Sigal, P. L. Darke, and J. P. Springer. 1989. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature (London)* **337**:615-620.
 42. Oroszlan, S., and R. B. Luftig. 1990. Retroviral proteinases. *Curr. Top. Microbiol. Immunol.* **157**:153-185.
 43. Palmiter, R. D., J. Gagnon, V. M. Vogt, S. Ripley, and R. N. Eisenman. 1978. The NH₂-terminal sequence of the avian oncovirus gag precursor polyprotein (Pr76^{gag}). *Virology* **91**:423-433.
 44. Pearl, L. H., and W. R. Taylor. 1987. Sequence specificity of retroviral proteases. *Nature (London)* **328**:482.
 45. Pearl, L. H., and W. R. Taylor. 1987. A structural model for the retroviral proteases. *Nature (London)* **329**:351-354.
 46. Peng, C., N. T. Chang, and T. W. Chang. 1991. Identification and characterization of human immunodeficiency virus type I gag-pol fusion protein in transfected mammalian cells. *J. Virol.* **65**:2751-2756.
 47. Peng, C., B. K. Ho, T. W. Chang, and N. T. Chang. 1989. Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity. *J. Virol.* **63**:2550-2556.
 48. Pepinsky, R. B. 1983. Localization of lipid-protein and protein-protein interactions within the murine retrovirus gag precursor by a novel peptide-mapping technique. *J. Biol. Chem.* **258**:11229-11235.
 49. Pettit, S. C., J. Simsic, D. D. Loeb, L. Everitt, C. A. Hutchison III, and R. Swanstrom. 1991. 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.
 50. Pettit, S. J., and R. Swanstrom. Personal communication.
 51. Poorman, R. A., A. G. Tomasselli, R. L. Heinrikson, and F. J. Kezdy. 1991. A cumulative specificity model for proteases from human immunodeficiency virus types 1 and 2, inferred from statistical analysis of an extended substrate data base. *J. Biol. Chem.* **266**:14554-14561.
 52. Power, M. D., P. A. Marx, M. L. Bryant, M. B. Gardner, P. J. Barr, and P. A. Luciw. 1986. Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. *Science* **231**:1567-1572.
 53. Skalka, A. M. 1989. Retroviral proteases: first glimpses at the anatomy of a processing machine. *Cell* **56**:911-913.
 54. Staden, R. 1988. Methods to define and locate patterns of motifs in sequences. *Comput. Appl. Biosci.* **4**:53-60.
 55. Stewart, L., G. Schatz, and V. M. Vogt. 1990. Properties of avian retrovirus particles defective in viral protease. *J. Virol.* **64**:5076-5092.
 56. Stewart, L., and V. M. Vogt. 1991. *trans*-acting viral protease is necessary and sufficient for activation of avian leukosis virus reverse transcriptase. *J. Virol.* **65**:6218-6231.
 57. Toh, H., R. Kikuno, H. Hayashida, T. Miyata, W. Kugimiya, S. Inouye, S. Yuki, and K. Saigo. 1985. Close structural resemblance between putative polymerase of a *Drosophila* transposable genetic element 17.6 and *pol* gene product of Moloney murine leukemia virus. *EMBO J.* **4**:1267-1272.
 58. Tritch, R. J., Y.-S. E. Cheng, F. H. Yin, and S. Erickson-Viitanen. 1991. Mutagenesis of protease cleavage sites in the human immunodeficiency virus type I gag polyprotein. *J. Virol.* **65**:922-930.
 59. Wlodawer, A., M. Miller, M. Jaskolski, B. K. Sathyanarayana, E. Baldwin, I. T. Weber, L. M. Selk, L. Clawson, J. Schneider, and S. B. Kent. 1989. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **245**:616-621.
 60. Yoshioka, K., H. Honma, M. Zushi, S. Kondo, S. Togashi, T. Miyake, and T. Shiba. 1990. Virus-like particle formation of *Drosophila copia* through autocatalytic processing. *EMBO J.* **9**:535-541.
 61. Youngren, S. D., J. D. Boeke, N. J. Sanders, and D. J. Garfinkel. 1988. Functional organization of the retrotransposon Ty from *Saccharomyces cerevisiae*: Ty protease is required for transposition. *Mol. Cell. Biol.* **8**:1421-1431.