JOURNAL OF VIROLOGY, Jan. 1991, p. 272–280
0022-538X/91/010272-09$02.00/0

Vol. 65, No. 1

Amino Acids Encoded Downstream of *gag* Are Not Required by Rous Sarcoma Virus Protease during Gag-Mediated Assembly

ROBERT P. BENNETT,[1] SUNG RHEE,[2] REBECCA C. CRAVEN,[1] ERIC HUNTER,[2] AND JOHN W. WILLS[1]*

Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center—Shreveport, 1501 Kings Highway, P.O. Box 33932, Shreveport, Louisiana 71130,[1] and Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294[2]

Received 8 August 1990/Accepted 10 October 1990

Rous sarcoma virus (RSV) and its relatives are unique in that they appear to encode their viral protease in the *gag* reading frame. As a result, this 124-amino-acid sequence is found at the carboxy terminus of each Gag precursor molecule and, upon ribosome frameshifting, embedded within each Gag-Pol molecule. However, rigorous proof has never been obtained for the activity of this 124-amino-acid Gag domain during virion assembly in vivo. If the active protease actually included amino acids encoded downstream in the *pol* reading frame, then the sequence organization would be more in line with those of other retroviruses. To examine this issue, mutations that disrupt the addition of amino acids by ribosome frameshifting were analyzed for their effects on particle assembly and Gag processing in a mammalian expression system (J. W. Wills, R. C. Craven, and J. A. Achacoso, J. Virol. 63:4331–4343, 1989). A 2-base substitution which created a nonsense mutation in the *pol* reading frame and was predicted to disrupt the hairpin structure of the ribosome frameshift signal had no effect on particle assembly or Gag processing, definitively showing that downstream amino acids are unnecessary. Mutations that fused the *gag* and *pol* reading frames to place 85 amino acids at the carboxy terminus of Gag hindered particle assembly and totally abolished the activity of the protease. A smaller fusion protein containing only the seven-amino-acid spacer peptide that links Gag and reverse transcriptase allowed particle formation but slowed processing. The reduced rate of processing exhibited by this mutant also revealed a previously unnoticed series of late maturation steps associated with the RSV capsid (CA) protein. Another mutant containing two substituted amino acids plus one additional amino acid at the carboxy terminus of protease nearly abolished processing. Together, these results demonstrate the importance of the carboxy terminus for proteolytic activity and suggest that this end must be unrestrained for optimal activity. If this hypothesis is correct, then the RSV protease may be encoded at the end of *gag* simply to ensure the production of a free carboxy terminus by translational termination.

The major internal structural proteins of all retroviruses are synthesized as polyprotein precursors which are proteolytically cleaved during virion morphogenesis. Translation of the *gag* and *pol* sequences in the unspliced viral RNA results in the synthesis of the Gag polyprotein precursor and, as a result of ribosomal frameshifting or suppression of a termination codon at the end of *gag*, the Gag-Pol precursor. The Gag protein drives particle formation at the plasma membrane (2, 24, 30), while the Gag-Pol protein provides the enzymatic activities (reverse transcriptase [RT] and integrase [IN]) needed for the subsequent synthesis and integration of the proviral DNA in the host cell (26). Because the frequency of frameshifting or termination suppression is low (5 to 10%) in all retroviruses (6), a great excess of Gag is produced relative to Gag-Pol. It is widely believed that the virus-encoded protease (PR) carries out all the cleavages on the Gag and Gag-Pol precursors and that these occur very late in, or after, the budding process (for recent reviews, see references 2, 18, and 23).

Many features of the retroviral proteases are highly conserved. For example, they are members of the aspartic acid protease family, and all require the formation of a dimer to assemble the active site (8, 19). Also, the crystal structures of two retroviral proteases, that of human immunodeficiency virus type 1 (HIV-1) and that of Rous sarcoma virus (RSV), have been determined and found to be very similar (13, 15, 16, 23, 28, 32). Even the general location of the PR-coding sequence, at a position between the nucleocapsid (NC)- and reverse transcriptase (RT)-coding sequences, is absolutely conserved.

In spite of these common features, there are some quite interesting differences between retroviral proteases. With regard to crystal structure, for example, the RSV protease contains extra loops that are not found in the structure of the smaller protease of HIV-1 (23, 32). Also, while all the other proteases contain the characteristic Asp-Thr-Gly sequence in the active site, Thr is replaced with Ser in the RSV protease (19). But perhaps the most striking differences of all are the amounts of protease that various retroviruses produce and the distribution of the protease domain among the polyprotein precursors. Sequence analysis of cloned retroviral genomes has shown that the PR-coding sequence may be located in either the *gag* reading frame, the *pol* reading frame, or a third frame, depending on the particular virus examined. In most viruses (e.g., HIV-1 and murine leukemia virus), PR is encoded at the 5′ end of the *pol* gene and is expressed only as part of the Gag-Pol fusion protein. In this situation, the protease is completely embedded within the polyprotein and therefore must be cleaved, presumably in an autocatalytic manner, to create both its amino and carboxy termini during release from the precursor. In the cases of mouse mammary tumor virus and human T-cell lymphotropic virus types I and II, the protease is encoded in the third reading frame between *gag* and *pol* such that PR is synthe-

* Corresponding author.

sized both as a Gag-PR precursor with a free carboxy terminus and embedded in a Gag-PR-Pol precursor (7).

In contrast to all other retroviruses, RSV and its relatives in the avian sarcoma-leukosis virus (ASLV) family encode a protease at the 3′ end of *gag*; therefore, this protease is found on the carboxy terminus of every Gag molecule and, by ribosomal frameshifting, embedded within every Gag-Pol molecule. While the protease in Gag is linked to the NC protein at its amino terminus, translation termination results in a free carboxy-terminal end (21). The protease in Gag-Pol is also linked to NC, but its carboxy terminus is linked to RT by means of a short (seven-amino-acid) spacer peptide (see Fig. 1; 22). As a consequence of the unique gene organization of the ASLV family, the protease of these viruses is produced in an equimolar amount with both of its substrates, the Gag and Gag-Pol precursors, and with the major core proteins (matrix [MA], p10, capsid [CA], and NC) that result from its activity on Gag.

It is not understood why RSV has evolved to have protease contained within each of its Gag and Gag-Pol molecules. It has been suggested that sequences encoded in the downstream *pol* reading frame might be required for activation of the protease during the process of particle formation (18). That is, it is possible that the protease located in the Gag precursor, though apparently active in vitro (27) and in *Escherichia coli* systems (11), is inactive in vivo. If this were the case, then the sequence organization of RSV would be more like that of all other retroviruses.

The experiments described in this report were designed to rigorously determine if the 124-amino-acid protease contained within the RSV Gag protein is sufficient for Gag processing during particle formation. By using oligonucleotide-directed mutagenesis to introduce stop codons and frameshift mutations at the *gag-pol* junction, a variety of Gag derivatives having differing amounts of *pol*-encoded amino acids at their carboxy termini were created. The mutants were analyzed in mammalian cells by using a previously described, simian virus 40 (SV40)-based, transient expression vector. For the particles that are produced in this system, their rate of release into the growth medium, their morphological appearance, and the electrophoretic mobilities of their mature cleavage products are all identical to those of authentic RSV (29, 30). The results presented here demonstrate that neither the spacer peptide nor portions of Pol are necessary for RSV protease activity in vivo and that sequences fused to the carboxy terminus of PR appear to be inhibitory to the processing pathway.

MATERIALS AND METHODS

DNAs and cells. The wild-type RSV *gag* gene was obtained from pATV-8, a molecular clone containing an infectious, sequenced copy of the RSV Prague C genome (22). Standard protocols were used for all of the DNA manipulations (20). MGAG, the recombinant M13 bacteriophage containing the *gag* gene and used for the mutagenesis, has been described previously (30). All phage stocks were routinely propagated in *E. coli* MV1190 (20) grown in LB medium. The SV40-based expression vector, pSV.Myr1, used to express the mutant *gag* genes in mammalian cells, has been described previously (30). All recombinant plasmids were propagated in *E. coli* DH-1 (20) with LB medium containing ampicillin (25 μg/ml). SV40-transformed African green monkey kidney cells (COS-1) were grown in Dulbecco's modified Eagle's medium (GIBCO Laboratories) supplemented with 3% fetal bovine serum and 7% calf bovine serum (HyClone, Inc.).

Oligonucleotide-directed mutagenesis. The uracil-substituted, single-stranded DNA template used for the mutagenesis reactions was isolated from MGAG grown in *E. coli* CJ236, a *dut ung* strain. Mutations were introduced into the *gag-pol* junction by using synthetic oligonucleotides and the method of Kunkel et al. (12). The oligonucleotide used to create SR12 was 5′-GCTTGACAAATTTTATAGGG-3′; that used for SR13 was 5′-CACTGTTCTCTAGGTTGCGC-3′; that used for SR14 was 5′-GCTTGACAAATTTAT AGGGTAGGCCACTGTTC-3′; and that used for SR15 was 5′-GCTTGACAAATTTAATAGGG-3′ (see Fig. 1). Mutant GS-1 was kindly provided by Volker M. Vogt (Cornell University, Ithaca, N.Y.). All the mutations were confirmed by DNA sequencing using the dideoxy-chain termination method.

Transfer of the mutations into the mammalian expression vector. DNA fragments containing the mutations were excised and gel purified from the double-stranded, replicative form MGAG DNAs by digestion with *Bgl*II, which cuts at nucleotide (nt) 1630 in the *gag* gene, and *Bss*HII, which cuts at nt 2724 in the *pol* gene. The SV40-based expression vector was also cut with the same two restriction endonucleases, and the larger of the two fragments (missing the *gag-pol* junction) was gel purified. The mutant fragments were inserted into the vector with T4 DNA ligase (New England BioLabs, Beverly, Mass.), and the recombinant DNAs were transformed into *E. coli* DH-1. To simplify screening for the desired clones, a derivative of pSV.Myr1, pSV.D37I, which contains an *Eco*RV site within the PR-coding sequence, was used (30). All the clones were screened for the loss of this site. The presence of the desired mutations in the recombinant clones was also confirmed by DNA sequencing.

Transfection of mammalian cells and metabolic labeling. COS-1 cells were transfected by the DEAE-dextran-chloroquine method, and approximately 48 h later, they were labeled with [^{35}S]methionine for 2.5 h as previously described (29–31).

Cell lysis, immunoprecipitation, and detection of Gag and Gag fusion proteins. The labeling medium was removed from the cells, microcentrifuged to remove loose cells, and mixed with detergents, as previously described (30). Cell lysates were prepared and the nuclei were discarded also as previously described (30). Gag-related proteins were immunoprecipitated from the lysate and medium fractions with an excess of rabbit antiserum raised against purified RSV (29). The antigen-antibody complexes were collected with fixed *Staphylococcus aureus* cells by standard protocols (5). All immunoprecipitated proteins were dissolved in sample buffer and separated in 1.5-mm-thick sodium dodecyl sulfate (SDS)–12% polyacrylamide gels as previously described (29, 30). Gels 10 cm in length were used for most experiments, but to achieve maximal resolution of the triplet of CA-related proteins, gels of 25 cm were required. The positions of labeled proteins in the gels were determined by fluorography using Fluoro-Hance (Research Products International, Inc.) and X-Omat AR5 film (Eastman Kodak Co., Rochester, N.Y.) at −70°C.

RESULTS

The RSV Gag protein, Pr76^{gag}, has been previously shown to contain a protease. The experiments described below were designed to test the hypothesis that amino acids encoded in the *pol* reading frame are essential for the activity of this protease during viral maturation. Stated another way, the goal of this study was to evaluate whether the protease

that is active in vivo is actually the product of ribosomal frameshifting at the end of *gag* rather than the more abundant protease that is synthesized as the terminal domain of Gag. To test this hypothesis, we have introduced stop codons in *pol* to cause premature termination of the Gag-Pol precursor. One of these mutations should also disrupt ribosomal frameshifting at the *gag-pol* junction. Further, we have constructed insertion mutations to align the *gag* and *pol* reading frames and thereby fuse either 7 or 85 amino acids of Pol onto Gag to evaluate the ability of such proteins to function in particle assembly and protease activity.

Mutagenesis strategy. Mutations were introduced by oligonucleotide-directed mutagenesis into the *Sac*I-*Hin*dIII fragment of the RSV Prague C genome carried in an M13 vector, as described in Materials and Methods. The mutations that were obtained are illustrated in Fig. 1. Restriction fragments containing the mutations were moved into a *gag* expression vector, pSV.Myr1, for evaluation of the protein products in COS-1 cells (see below). The amino acid sequences that are expected to result from the expression of the different mutants are also shown in Fig. 1.

Two mutations were created by the introduction of stop codons in the Gag-Pol spacer region to cause premature termination of the *gag-pol* readthrough product. The first truncation mutant, designated SR13, was made by changing the eighth codon of the Pol reading frame from ACT (Thr) to TAG (stop). Because this 3-nt substitution falls within the loop of the frameshifting signal, as characterized by Jacks et al. (6), we expected that this mutation might have no effect on the efficiency of ribosomal frameshifting. Thus, translation with normal frameshifting at the end of *gag* should produce $Pr76^{gag}$ as the predominant product, with small amounts (approximately 5%) of a readthrough product consisting of the $Pr76^{gag}$ fused to the seven-amino-acid spacer. Because of the small difference in size (seven amino acids), we were not able to reliably resolve these two products (Fig. 2). The second truncation mutant, SR14, was created by changing the third codon in the *pol* reading frame from AGG (Arg) to TAG (stop). Frameshifting was expected to be disrupted by this mutation, since the two nucleotides affected are involved in base pairing within the stem of the ribosome frameshift signal (6). However, if frameshifting were to continue, then a small fraction of Gag molecules synthesized should consist of $Pr76^{gag}$ fused to the first two amino acids of the spacer peptide (Fig. 1).

Two (+1) frameshift mutations were generated by the insertion of single nucleotides at the *gag-pol* junction to evaluate whether the Gag protease would be proteolytically active when amino acids derived from the spacer peptide and Pol were fused to the carboxy terminus of PR. The first of these mutants, SR12, was produced by a thymidine insertion in the final *gag* codon. SR15 was created by the insertion of an adenine immediately preceding the *gag* termination codon (Fig. 1). Both changes should have the effect of placing the *gag* stop codon out of frame and result in the fusion of the *gag* and *pol* reading frames. Translation is expected to produce a protein consisting of $Pr76^{gag}$ fused to 86 amino acids of the spacer and Pol, with a single amino acid substitution at the end of the protease sequence in the case of SR12. No $Pr76^{gag}$ should be produced by either mutant. If frameshifting still occurred, then a minor translation product composed of $Pr76^{gag}$ plus 28 foreign amino acids encoded by the third reading frame would be expected. However, the frequency with which this species should arise cannot be predicted. Though Jacks et al. (6) have shown that UUU at the site of frameshifting actually increases the efficiency of frameshifting, in SR12 and SR15, the distance between the stem-loop of the frameshifting signal and the actual site of frameshifting has been increased by 1 nt, making it unclear what effect this will have on the frequency of the frameshift event.

Gag-Pol truncation mutants SR13 and SR14. The effects of the SR13 and SR14 mutations on particle formation and proteolytic processing were analyzed by transfection of plasmid DNA containing these alterations into COS-1 cells as described previously (29, 30). At 48 h posttransfection, the cells were labeled with [^{35}S]methionine for 2.5 h. Gag proteins from the media and cell lysates were immunoprecipitated with an anti-RSV serum, separated by electrophoresis in an SDS–12% polyacrylamide gel, and visualized by fluorography. In this type of assay, assembly-competent Gag proteins, such as the Myr1 protein expressed by the parent plasmid, pSV.Myr1, rapidly form particles that are released from the cell (29, 30), as evidenced by the accumulation of Gag proteins in the labeling medium (Fig. 2A, Media lane 6). When the Gag protein carries an active PR, as is the case with Myr1, processing to release the mature proteins is very rapid, and generally no uncleaved precursor can be seen in the medium. Rather, the mature cleavage products CA (p27) and PR (p15) are evident. Under these labeling conditions, the MA protein appears as a 23-kDa species, composed of MA and a small peptide named p2 (29, 30), which runs slightly faster than the CA bands.

When SR13 and SR14 were expressed in COS-1 cells, particle formation appeared to be equally efficient to that of Myr1, since the amounts of Gag products found in the medium samples are similar (Fig. 2A, lanes 3 and 5, and Fig. 2B, lanes 2 and 3). That these proteins were contained in viruslike particles was confirmed by sedimentation analysis through sucrose (data not shown). Furthermore, the electrophoretic mobilities of the mature Gag products present in the medium of SR13- and SR14-transfected cells appeared virtually identical to those of Myr1; processing was complete and only the proteolytic products were present. These results strongly suggest that the translation of a *gag-pol* readthrough product is not necessary for the function of the RSV *gag*-encoded protease. SR13 clearly demonstrates that nothing downstream of the spacer peptide is essential. In the case of SR14, if any amino acids of the spacer peptide are fused to PR, which seems unlikely, the fusion must be limited to two amino acids in a small fraction of the total translation products. Thus, it appears that the 124-amino-acid PR encoded by *gag* is sufficient for catalyzing the maturation of the Gag precursor in vivo.

Gag-Pol fusion mutants SR12 and SR15. These two mutants were designed to determine whether the protease could still function if the major translation product were a fusion protein consisting of Gag fused to the first 85 amino acids encoded in the *pol* reading frame. Upon expression of SR12 and SR15, no proteins of a length corresponding to $Pr76^{gag}$ were observed in either the lysates or media, as predicted (Fig. 2A, lane 2, and Fig. 2C, lanes 2 and 3). Instead, a protein product of the size expected for the fusion product was present. The behavior of this protein for each of the two mutants was identical. Large amounts of the uncleaved precursor accumulated in the cells, but very little was released into the medium and no cleavage products were observed in the medium. In a pulse-chase labeling experiment with mutant SR12, no processing of this polyprotein was observed even after a 6-h chase (Fig. 3). Thus, it is clear that fusion of even a small portion of Pol (85 amino acids) to the carboxy terminus of the RSV protease hinders particle

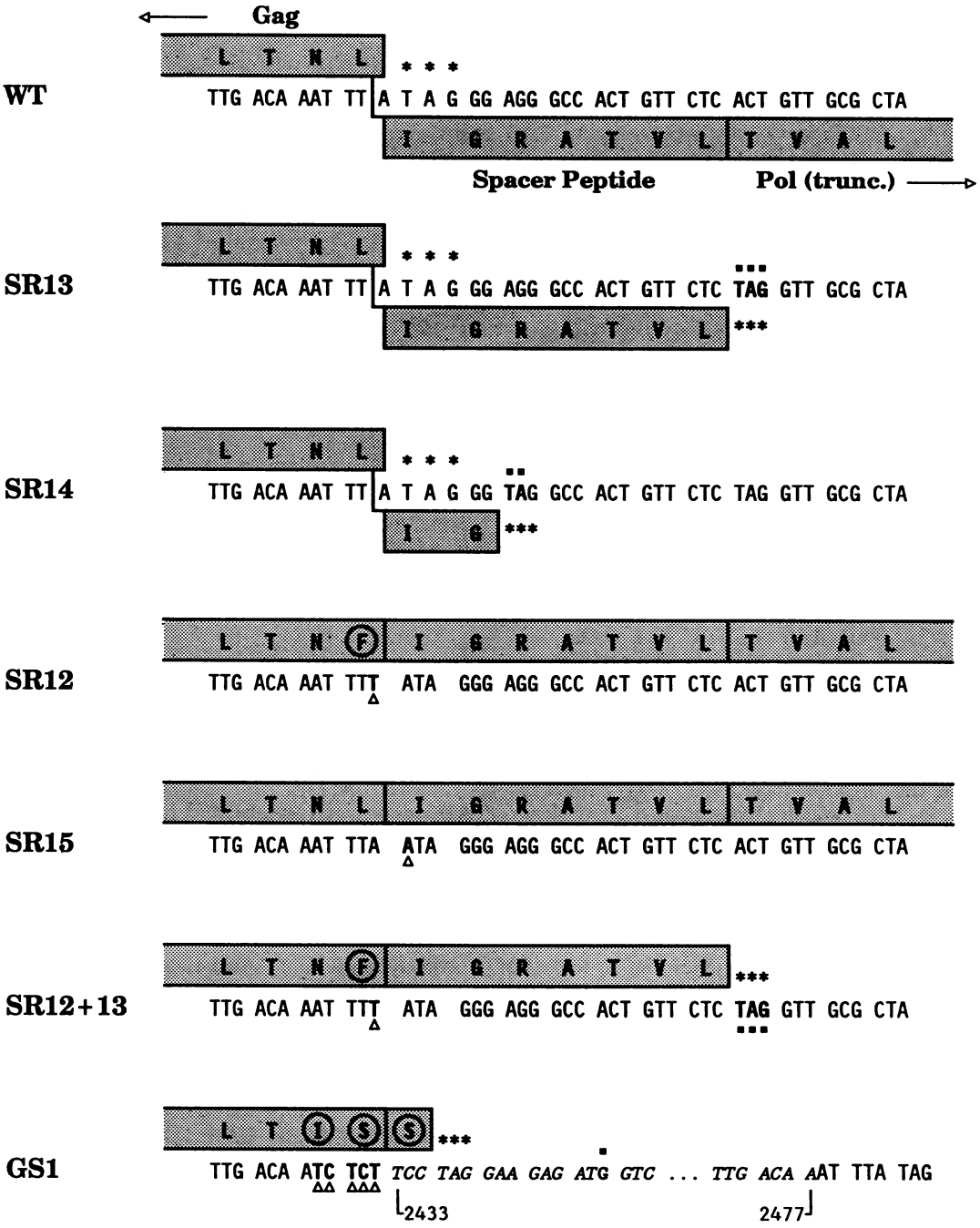

FIG. 1. Mutagenesis of the *gag-pol* spacer region. The nucleotide sequence of the *gag-pol* junction for the wild type (WT) and each of the mutants is given. The shaded boxes contain the predicted *gag* and *pol* products, listed above and below the nucleotide sequence, respectively. Open-ended boxes indicate that the amino acid sequence extends to the right or left, as described in the text; ***, stop codons. The nucleotide substitutions are indicated by solid squares either above or below the altered sequence. The open arrowheads denote inserted nucleotides. Circled amino acids indicate substitutions. Mutant GS1 contains a 5-nt insertion (open arrowheads) and a duplication of an upstream sequence (nt 2433 through 2477; italicized and marked with brackets). A single point mutation within the duplication is indicated by a solid square.

formation dramatically when this protein is expressed as the major translation product. Further, in those particles that are produced, the protease is completely inactive.

We next sought to determine whether the fusion of a shorter peptide, the seven amino acids of the spacer, onto the carboxy terminus of Gag would have a similar inhibitory effect on particle formation and protease activity. This was accomplished by combining the SR12 and SR13 mutations to create SR12+13 (Fig. 1). The behavior of this protein in COS-1 cells is shown in Fig. 2A (lanes 4) and Fig. 3. The Gag precursor seen in the lysate of SR12+13-transfected cells is detectably larger in size than the precursor expressed by Myr1, SR13, and SR14, consistent with its predicted structure. The SR12+13 Gag protein, unlike either SR12 or SR15,

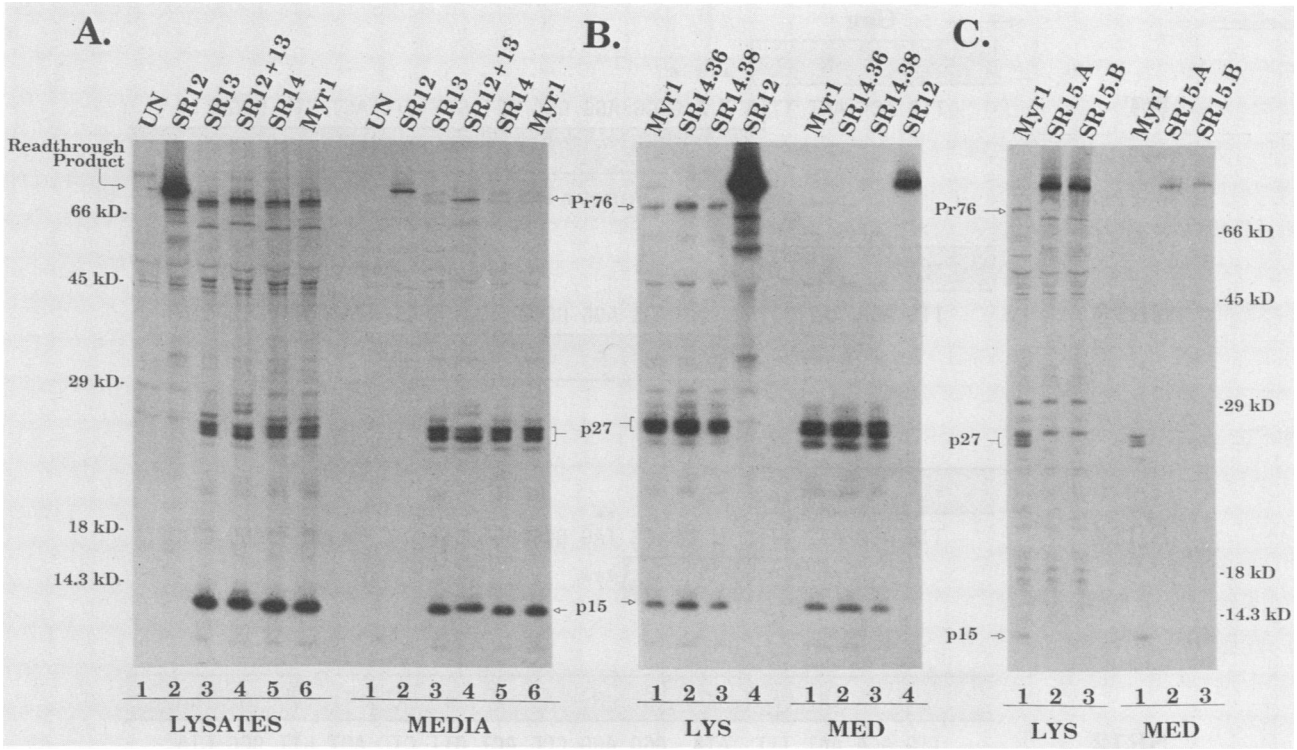

FIG. 2. Expression of Gag and truncated Gag-Pol proteins in COS-1 cells. COS-1 cells transfected with the Myr1 parent or the various mutant DNAs were labeled with [^{35}S]methionine for 2.5 h, and then the RSV-related proteins present in the cell lysates and in the media were collected by immunoprecipitation using an anti-RSV serum as detailed in Materials and Methods. The precipitated proteins were electrophoresed in an SDS–12% polyacrylamide gel and visualized by fluorography. (A) Proteins produced by mutants SR12, SR13, SR14, and SR12+13 are compared with those from the Myr1 parent and untransfected (UN) cells. (B) Comparison of the products of two clones of SR14 with those of SR12 and the Myr1 parent. (C) Two clones of SR15 are compared with the Myr1 parent. The position of the *gag-pol* readthrough product of Myr1 is indicated by the arrow at the left. This position also corresponds to the SR12 and SR15 fusion proteins. The positions of Gag (Pr76), the CA triplet (p27), and PR (p15) are indicated. The positions of protein size standards within the gel are shown at the right and left of the figure.

is able to form particles with an efficiency similar to that of Myr1, as indicated by the large amount of Gag products that were seen in the medium. However, the efficiency of proteolytic processing seemed to be reduced, since a significant amount of uncleaved Gag precursor was released into the medium (Fig. 2A, Media lane 4). The pulse-chase labeling experiment whose results are shown in Fig. 3A confirmed this effect. After 15 min of pulse labeling, cleavage products from the wild-type protein (Myr1) were already apparent, whereas the SR12+13 product remained almost entirely in the precursor form in the lysate and medium samples. Though the rate of processing of the SR12+13 protein was reproducibly slower than that of Myr1, within 2 h it appeared to be nearly complete.

We also found that mutant SR12+13 consistently produces readily detectable levels of a processing intermediate that migrates slightly slower than mature PR (Fig. 3B, lane 2), and this species probably corresponds to the protease with the spacer still attached, though definitive proof of this awaits further experimentation. As expected, this band was not found with SR14 (lane 3), but it is occasionally detected in trace amounts in SR13 (lane 1) and the parent Myr1 (not shown). In pulse-chase experiments (Fig. 3A), the band was visible after 2 h of chase but disappeared by 6 h to yield mature PR. Since each of the SR12+13 Gag precursors is synthesized with the spacer peptide fused to it, the appearance of a cleavage product of a size similar to that of the wild-type PR implies that cleavage to remove the spacer has occurred in spite of the amino acid substitution (Leu to Phe) at the end of PR. This substitution may account for the very slightly slower migration of the mature protease from SR12+13 relative to SR13 and SR14 (Fig. 3B).

CA maturation. In addition to the slower rate of Gag processing, a subtle difference was noted in the electrophoretic profile of the CA (p27) protein produced by mutant SR12+13 relative to the wild-type products. Under our standard conditions for [^{35}S]methionine labeling and gel electrophoresis, the wild-type CA protein is generally resolved into a set of three closely migrating bands of approximately equal intensity (Fig. 3C, lane 4). All three of these bands reacted with antibodies raised against two different synthetic peptides from CA, indicating that all are forms of capsid protein (data not shown). This wild-type CA pattern was also observed with mutants SR13 (lane 1) and SR14 (lane 3), but in the case of mutant SR12+13, the fastest-migrating band of the triplet was more intense (lane 2). Because the protease of mutant SR12+13 is impaired in its overall rate of processing, it seemed reasonable that one of the CA species might be a precursor from which the others are derived by proteolysis.

To further explore the maturation steps that lead to the appearance of the three CA species, a pulse-chase labeling experiment was performed using the wild-type (Myr1) protein. Three identical plates of pSV.Myr1-transfected COS-1

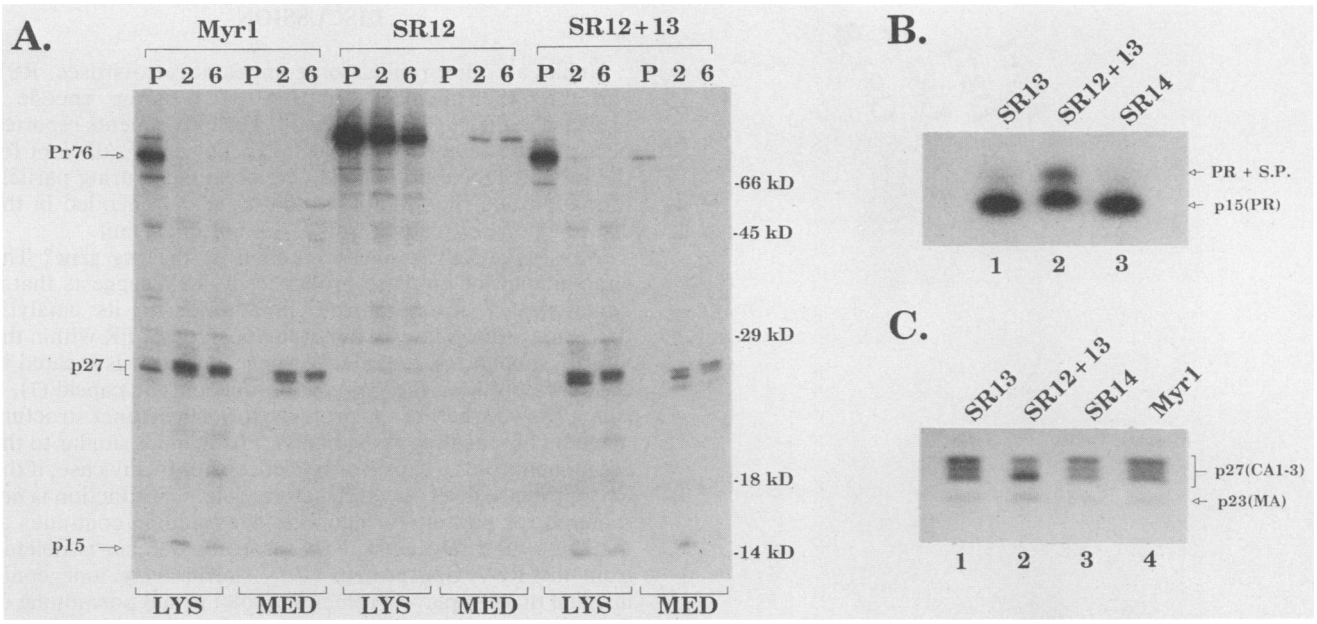

FIG. 3. Kinetics of wild-type and mutant Gag precursor processing. COS-1 cells were transfected with the DNAs indicated across the top of each panel and labeled with [^{35}S]methionine 48 h later. (A) Pulse-chase labeling of mutant proteins. After being pulse-labeled for 15 min, the cells were incubated in medium containing an excess of unlabeled methionine for 0, 2, or 6 h (lanes P, 2, and 6, respectively). Gag-related proteins were collected by immunoprecipitation from cell lysate (LYS) and medium (MED) samples and electrophoresed in an SDS-polyacrylamide gel. The resulting fluorogram is shown. Positions of protein size standards within the gel are indicated at the right. Positions of the wild-type RSV *gag* products are indicated at the left. (B) Comparison of mutant protease species. Transfected cells were labeled for 2.5 h prior to immunoprecipitation. Only that portion of the resulting fluorogram corresponding to the PR (p15) proteins is shown. S.P., Spacer peptides. (C) Altered pattern of CA maturation for mutant SR12+13. Transfected cells were labeled for 2.5 h before immunoprecipitation. Only that portion of the resulting fluorogram corresponding to the CA (p27) and MA (p23) proteins is shown.

cells were labeled with [^{35}S]methionine for 15 min and then chased in medium with unlabeled methionine for various times. To best resolve the CA bands, proteins immunoprecipitated with anti-RSV serum were electrophoresed in a long (25-cm) SDS-polyacrylamide gel (12% acrylamide). After a 15-min chase, only the fastest-migrating CA species was detected in the lysate and medium (Fig. 4). After a 60-min chase, the other two CA species became apparent, but at a level that was reduced relative to the initial species. This unequal distribution is reminiscent of the altered pattern of mutant SR12+13 (Fig. 3C). After a 3-h chase, all three species of CA were evident in equal proportions (Fig. 4). With still longer chase times (6 h), only the upper two CA species remained, even in the case of mutant SR12+13 (Fig. 3A).

From these experiments, we conclude that CA maturation initially involves a very rapid cleavage event which, surprisingly, produces the fastest-migrating species, designated CA1, and this is followed by a very slow series of events in which the two slower-migrating species, CA2 and CA3, arise. A determination of the exact nature of CA1, CA2, and CA3 awaits further investigation, but the data presented here suggest the hypothesis that CA2 and CA3 are proteolytically derived from CA1 since a mutation (SR12+13) that slows protease activity also slows the maturation of CA (see Discussion).

Alteration of the carboxy terminus of PR. Another Gag mutant, GS1, with an altered carboxy-terminal end was evaluated for protease activity during particle formation in COS-1 cells. The relevant nucleotide sequence for this mutant, which was a serendipitous result of an in vitro mutagenesis experiment, is shown in Fig. 1. The PR-coding sequence is wild type through nt 2477, after which a 5-nt insertion occurs, followed by an imperfect duplication of an upstream sequence (nt 2433 through 2477). There is a single point mutation (T to G at nt 2447) within this duplication. The predicted amino acid sequence indicates that GS1

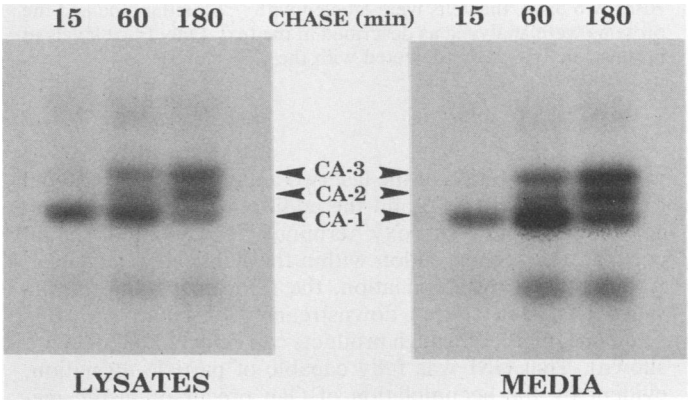

FIG. 4. Late maturation of the wild-type CA protein. Myr1-transfected cells were labeled with [^{35}S]methionine for 15 min, transferred to chase medium, and incubated for 15, 60, or 180 min as indicated. At those times, cell lysate and medium samples were prepared and the Gag-related proteins were immunoprecipitated with an anti-RSV serum. The collected proteins were analyzed on a 25-cm SDS-polyacrylamide gel. Only that portion of the resulting fluorogram corresponding to the CA proteins is shown. The band running below the CA triplet corresponds to p23, an MA-related species described in the text.

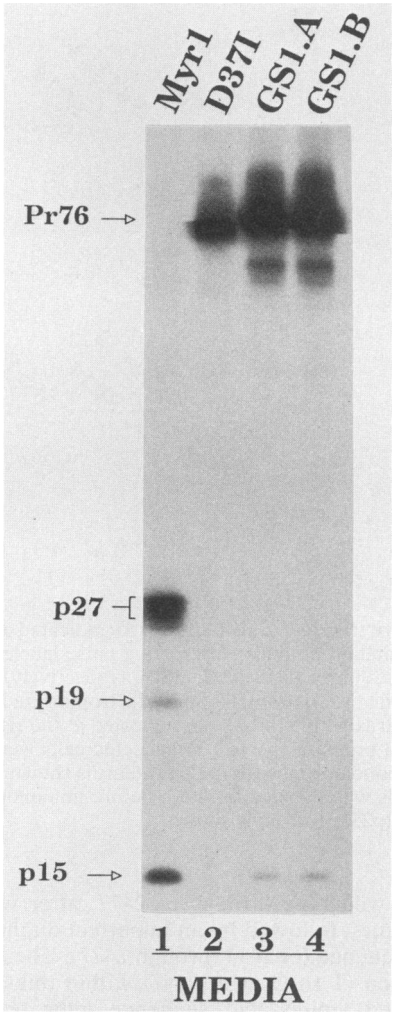

FIG. 5. Alteration of the carboxy terminus of PR abolishes protease activity. COS-1 cells were transfected with Myr1, Myr1 encoding a mutant protease (D37I), or one of two clones of GS1. After two days, the cells were labeled with [^{35}S]methionine and the proteins were analyzed as described in the text. Only trace levels of protease activity were detected with the GS1 clones.

carries a substitution of the last two amino acids of Gag (Asp to Ile and Leu to Ser) and the addition of a foreign amino acid (Ser) at the carboxy terminus. Termination should occur at the second codon within the duplicated sequence. As a result of the duplication, the ribosomal frameshifting signal has been shifted downstream. Consistent with this sequence, no readthrough products can be detected (data not shown). That GS1 was fully capable of particle formation, evident by the accumulation of Gag precursor in the medium, is shown in Fig. 5. Processing, however, appeared to occur very inefficiently, since only a small amount of PR was observed. Longer exposures revealed trace amounts of the Gag cleavage products, but after a long chase, additional processing did not appear to take place (data not shown). These results, together with those obtained with mutants SR12 and SR12+13 and described above, demonstrate the importance of a free and unmodified carboxy terminus for correct proteolytic activity in vivo.

DISCUSSION

Unlike all other replication-competent retroviruses, RSV and the other members of the ASLV group encode a protease in the *gag* reading frame. The experiments reported here clearly demonstrate that this protease is sufficient for the in vivo processing of the Gag precursor during particle formation and that additional amino acids, encoded in the downstream *pol* reading frame, are not important.

Why is the RSV protease encoded by the *gag* gene? The large amount of protease produced by RSV suggests that it could play a structural role in addition to its catalytic function. Little is known about the position of PR within the virion. Though some evidence suggests that it is located in the space between the viral membrane and the capsid (1), it is not known whether the protease forms a distinct structure made up of repeating PR subunits, in a manner similar to the components of the capsid or nucleocapsid. In any case, if the RSV protease does play a structural role, that function is not required for particle formation, since budding continues at the same high rate even if the protease domain is deleted from the RSV Gag protein (29). Alternatively, one could imagine that PR plays a role in the folding and positioning of the cleavage sites within the Gag precursor so that they remain accessible inside the environment of the immature virion. Experiments to begin addressing this idea would require that the RSV protease be packaged into particles at lower concentrations, perhaps by moving its coding sequence into the *pol* frame.

Reduced levels of RSV protease might be expected to impede Gag processing because this enzyme appears to have an inherently lower level of activity than other (non-ASLV) proteases (10). Thus, the RSV protease might be needed in large amounts simply to facilitate complete processing. Interestingly, sequence comparison of all known retroviral and cellular aspartic acid proteases has revealed that each contains the sequence Asp-Thr-Gly in its active site, except the ASLV protease, in which Thr is replaced with Ser (19). (The only other sequence in which Ser is found at this position is in the Gag-like protein of hamster intracisternal A particles [17].) It would be quite interesting to determine whether the RSV protease becomes more active, and capable of cleaving Gag at lower concentrations, if Ser-38 is changed to Thr.

Another possible reason why the RSV protease might be needed on every Gag molecule is that one or more cleavages might occur only in *cis*. That is, any two precursors that interact to form a dimeric protease might be cleaved intramolecularly only at certain positions by that same protease. However, in vitro cleavage of the Gag precursor (27) or peptide substrates (9–11) by the RSV protease has been demonstrated; thus, PR is not limited to a *cis* cleavage ability. Moreover, we have recently obtained evidence that shows that the RSV protease is capable of cleaving in *trans* during particle formation in complementation experiments (to be presented elsewhere).

The hypothesis that we presently favor for why the RSV protease is encoded by *gag* is that this position ensures that a free carboxy-terminal end will be created by translational termination. This appears to be important because the activity of the protease is greatly impeded or eliminated if amino acids are attached to its carboxy terminus. This was shown by mutants SR12 and SR15, in which the addition of 85 *pol*-encoded amino acids to the end of the Gag precursor completely halted processing. Because this polypeptide fragment from RT is unlikely to have the capacity to fold

properly, one could argue that it is likely to interfere with the proper folding of the immediately adjacent PR domain. However, even when just the seven-amino-acid spacer peptide was attached to the Gag protein (in mutant SR12+13), there appeared to be an inhibition of the protease, though we have not ruled out the possibility that the substitution of Leu with Phe at residue 124 accounted for the reduced activity. If our free-end hypothesis is correct, then the protease embedded within the Gag-Pol precursor would be expected to have little or no activity and thus to be unable to generate its amino or carboxy terminus by an autocatalytic cleavage event. Experiments to test this idea are in progress using our mammalian expression system.

Whether or not the free-end hypothesis is correct, it is clear that the amino acids at the very amino terminus of the ASLV protease are critical for activity. Minor alteration of this sequence in mutant GS-1 almost completely abolished protease activity. Similar results have been observed with the HIV-1 protease (4, 14). The crystal structures of the RSV and HIV-1 proteases are superimposable in this region and suggest that the carboxy termini, which form antiparallel β strands with the amino termini, are critical for dimerization (13, 15, 32).

Why do additional amino acids interfere with particle formation? Mutations SR12 and SR15 had a profound inhibitory effect on particle formation. Both of these resulted in the addition of 85 residues of Pol to the carboxy terminus of Gag, creating fusion proteins of 786 amino acids. It is not clear what accounts for the interference in particle formation. It may be that the additional amino acids cannot be accommodated easily within the limited internal space of the particle. This seems unlikely because in our laboratory iso-1-cytochrome *c* from *Saccharomyces cerevisiae* has been fused to the carboxy-terminal region of the RSV Gag protein to produce a chimeric protein that is only 31 amino acids shorter than the SR12 and SR15 proteins, and this Gag fusion protein is very efficient at making particles (29). Another possibility, mentioned above in regard to the lack of processing of SR12 and SR15, is that the addition of an improperly folded, truncated segment of the Pol domain might inhibit the folding of Gag and consequently particle formation. Though we cannot rule this possibility out, we point out that in the Gag-cytochrome fusion, the foreign sequence was joined at a position internal to the protease domain and the remaining, truncated portion of protease did not interfere. Moreover, we have recently fused the RSV *gag* and *pol* reading frames to express high levels of the full-length Gag-Pol protein in the absence of Gag-only molecules and found that these are also incapable of particle formation (to be presented elsewhere); identical results have been obtained with the Gag-Pol protein of murine leukemia virus (3). A third possibility is that the sequence encoded immediately downstream of *gag* contains a negative regulatory domain that prevents Gag-Pol molecules from driving particle formation in the absence of Gag-only molecules. Though there is no precedent for this sort of a negative regulatory function, it would seem advantageous to regulate the maximal amounts of RT and IN that are packaged into a virion.

What is the role of the spacer peptide region? It appears that the spacer peptide does not play a direct role in the activity or regulation of the RSV protease. It is also irrelevant for particle formation. However, the coding sequence for this peptide coincides with the 5′ side of the stem region of the ribosome frameshift signal (6). Mutations made in this region that destabilize base pairing in the stem reduce frameshifting, and compensatory mutations that restore base pairing reinstate frameshifting. Thus, it seems most likely that the RNA sequence in this region is much more important than the translational product. Nevertheless, activation of the RSV reverse transcriptase appears to require its release from the Gag-Pol precursor (25), and the spacer peptide could play a role in coordinating the recognition of cleavage sites by the viral protease.

It is not known whether the spacer peptide is cleanly and completely excised from the Gag-Pol protein. Although the mature amino terminus of RT has been determined by amino acid sequencing, removal of the spacer peptide from the carboxy terminus of the protease has never before been demonstrated. The virus might not need to completely remove the spacer, since only 5% of the protease molecules (i.e., those derived from the Gag-Pol precursor) would ever be linked to it. In this report, we have described a Gag mutant, SR12+13, in which the spacer peptide is present on the end of every Gag molecule. Though this sequence slows proteolysis, it apparently does not block the removal of the spacer, even though the Leu residue normally found at the very end of PR (and in the P1 position of the cleavage site) has been substituted with Phe. Because the presence of an aromatic amino acid in the P1 position is common to many retroviral cleavage sites, it may not be surprising that cleavage occurs with SR12+13. The possibility that cleavage occurs in the wild-type situation, where Leu is present in this position, seems likely and could be determined by combining mutants SR15 and SR13.

Late maturation events associated with the RSV capsid protein. In a previous report, we noted that the RSV capsid protein ran as a doublet of bands in our SDS-polyacrylamide gels (30). Since then, we have slightly altered the composition of our gels (29) and have been able to resolve these bands into a triplet, as shown in this paper. This triplet is not an artifact of our mammalian expression system because it can also be detected by using authentic RSV grown in avian cells (data not shown). The pulse-chase experiments reported here demonstrate the maturation of CA1 to yield the two slower-migrating species, CA2 and CA3. If newly released particles containing a predominance of CA1 are removed from the cell monolayer and incubated further at 37°C, then CA maturation still continues, indicating that it is a postbudding event (data not shown).

The physical basis for capsid maturation is not certain, but we have noticed that mutant SR12+13 yields an increased amount of CA1 under our standard labeling conditions. Since this product eventually does give rise to CA2 and CA3, but at a rate that is consistent with the overall reduced efficiency of Gag processing for this mutant, it is likely that maturation involves proteolysis. Though it is conceivable that the maturation of CA1 results from modification by a cellular enzyme packaged within the virion, it is difficult to explain how such a cellular enzyme would be affected by a mutation in the viral protease. Because no other types of RSV capsid modification have been described, we presently favor the idea that CA2 and CA3 arise by the trimming of amino acids from the end(s) of CA1. It should be noted that a 9-amino-acid spacer peptide is located between the CA and NC sequences in the RSV Gag protein. It is possible that the species of capsid that includes this peptide corresponds to CA1 and that CA2 and CA3 lack some or all of the spacer. The reason why the mature products migrate more slowly than CA1 might have to do with the very hydrophobic nature of this spacer peptide. For example, if CA1 indeed contains this spacer and it binds a disproportionate amount of SDS

because of its hydrophobicity, then the polypeptide would run faster than expected in an SDS-polyacrylamide gel. Peptide mapping experiments are in progress to test these hypotheses about the physical basis of RSV capsid maturation.

ACKNOWLEDGMENTS

We thank Volker M. Vogt and Gisela Schatz for providing mutant GS-1 and Lance Stewart for reminding us that release of the spacer peptide from the carboxy terminus of the RSV protease had never before been demonstrated. We thank Moshe Kotler for reminding us that the ASLV protease contains Ser rather than Thr in the active site. We also thank Robert Weldon and Timothy Nelle for helpful discussions and review of the manuscript.

This work was supported by Public Health Service grants R29 CA47482 to J.W.W. and R01 CA27834 and R37 CA29884 to E.H. from the National Institutes of Health.

ADDENDUM IN PROOF

We have now constructed mutant SR15+13. Like mutant SR12+13, it exhibits a reduced rate of appearance of all the mature Gag cleavage products. It also appears to cleave the spacer from the carboxy-terminal end of PR, but at a rate that is even slower than for SR12+13. These results rule out the Leu to Phe substitution of SR12+13 as the cause for the reduced protease activity and further support our free end hypothesis.

REFERENCES

1. **Bolognesi, D. P., R. Luftig, and J. H. Sharper.** 1973. Localization of RNA tumor virus polypeptides. I. Isolation of further virus substructures. Virology **56**:549–564.
2. **Dickson, C., R. Eisenman, H. Fan, E. Hunter, and N. Teich.** 1984. Protein biosynthesis and assembly, p. 513–648. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
3. **Felsenstein, K. M., and S. P. Goff.** 1988. Expression of the *gag-pol* fusion protein of Moloney murine leukemia virus without *gag* protein does not induce virion formation or proteolytic processing. J. Virol. **62**:2179–2182.
4. **Hansen, J., S. Billich, T. Schulze, S. Sukrow, and K. Moelling.** 1988. Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. EMBO J. **7**:1785–1791.
5. **Harlow, E., and D. Lane.** 1988. Antibodies: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
6. **Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus.** 1988. Signal for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. Cell **55**:447–458.
7. **Jacks, T., K. Townsley, H. E. Varmus, and J. Majors.** 1987. Two efficient ribosomal frameshift events are required for synthesis of mouse mammary tumor virus *gag*-related polyproteins. Proc. Natl. Acad. Sci. USA **84**:4298–4302.
8. **Katoh, I., Y. Ikawa, and Y. Yoshinaka.** 1989. Retrovirus protease characterized as a dimeric aspartic proteinase. J. Virol. **63**:2226–2232.
9. **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.
10. **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.
11. **Kotler, M., R. A. Katz, and A. M. Skalka.** 1988. Activity of avian retroviral protease expressed in *Escherichia coli*. J. Virol. **62**:2696–2700.
12. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154**:367–382.
13. **Lapatto, R., T. Blundell, A. Hemmings, J. Overington, A. Wilderspin, S. Wood, J. R. Merson, P. J. Whittle, D. E. Danley, K. F. Geoghegan, S. J. Hawrylik, S. E. Lee, K. G. Scheld, and P. M. Hobart.** 1989. X-ray analysis of HIV-1 proteinase at 2.7 angstrom resolution confirms structural homology among retroviral enzymes. Nature (London) **342**:299–302.
14. **Loeb, D. D., R. Swanstrom, L. Everitt, M. Manchester, S. E. Stamper, and C. A. Hutchison.** 1989. Complete mutagenesis of the HIV-1 protease. Nature (London) **340**:397–400.
15. **Miller, M., M. Jaskolski, J. K. M. Rao, J. Leis, and A. Wlodawer.** 1989. Crystal structure of a retroviral protease proves relationship to aspartic protease family. Nature (London) **337**:576–579.
16. **Navia, M. A., P. M. D. Fitzgerald, B. M. McKeever, C. 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.
17. **Ono, M., H. Toh, T. Miyata, and T. Awaya.** 1985. Nucleotide sequence of the Syrian hamster intracisternal A-particle gene: close evolutionary relationship of type A particle gene to types B and D oncovirus genes. J. Virol. **55**:387–394.
18. **Oroszlan, S., and R. B. Luftig.** 1990. Retroviral proteinases. Curr. Top. Microbiol. Immunol. **157**:153–185.
19. **Pearl, L. H., and W. R. Taylor.** 1987. A structural model for the retroviral proteases. Nature (London) **329**:351–354.
20. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Sauer, R. T., D. W. Allen, and H. D. Niall.** 1981. Amino acid sequence of p15 from avian myeloblastosis virus complex. Biochemistry **20**:3784–3791.
22. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
23. **Skalka, A. M.** 1989. Retroviral proteases: first glimpses at the anatomy of a processing machine. Cell **56**:911–913.
24. **Smith, A. J., M.-I. Cho, M.-L. Hammarskjöld, and D. Rekosh.** 1990. Human immunodeficiency virus type 1 $Pr55^{gag}$ and $Pr160^{gag\text{-}pol}$ expressed from a simian virus 40 late replacement vector are efficiently processed and assembled into viruslike particles. J. Virol. **64**:2743–2750.
25. **Stewart, L., G. Schatz, and V. M. Vogt.** 1990. Properties of avian retrovirus particles defective in viral protease. J. Virol. **64**:5076–5092.
26. **Varmus, H., and R. Swanstrom.** 1984. Replication of retroviruses, p. 369–512. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Vogt, V. M., A. Wight, and R. Eisenman.** 1979. In vitro cleavage of avian retrovirus Gag proteins by viral protease p15. Virology **98**:154–167.
28. **Weber, I. T., M. Miller, M. Jaskolski, J. Leis, A. M. Skalka, and A. Wlodawer.** 1989. Molecular modeling of the HIV-1 protease and its substrate binding site. Science **243**:928–931.
29. **Weldon, R. A., C. R. Erdie, M. G. Oliver, and J. W. Wills.** 1990. Incorporation of chimeric Gag protein into retroviral particles. J. Virol. **64**:4169–4179.
30. **Wills, J. W., R. C. Craven, and J. A. Achacoso.** 1989. Creation and expression of myristylated forms of Rous sarcoma virus Gag protein in mammalian cells. J. Virol. **63**:4331–4343.
31. **Wills, J. W., R. V. Srinivas, and E. Hunter.** 1984. Mutations of the Rous sarcoma virus *env* gene that affect the transport and subcellular localization of the glycoprotein product. J. Cell Biol. **99**:2011–2023.
32. **Wlodawer, A., M. Miller, M. Jaskolski, B. K. Sathyanarayana, E. Baldwin, I. T. Weber, L. M. Selk, L. Clawson, J. Schneider, and S. B. H. Kent.** 1989. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science **245**:616–621.