Maturation of Giardavirus Capsid Protein Involves Posttranslational Proteolytic Processing by a Cysteine Protease

DECHAO YU, CHING C. WANG, AND ALICE L. WANG*

Department of Pharmaceutical Chemistry, University of California,
San Francisco, California 94143

Received 30 November 1994/Accepted 2 February 1995

The double-stranded RNA genome of giardavirus (GLV) has only two large open reading frames (ORFs). The 100-kDa capsid polypeptide (p100) is encoded by ORF1, whereas the only other viral polypeptide, the 190-kDa GLV RNA-dependent RNA polymerase (p190), is synthesized as an ORF1-ORF2 fusion protein by a (–1) ribosomal frameshifting. Edman degradation revealed that p100 was N-terminally blocked except for 2 to 5% of it that showed free N terminus starting from amino acid residue 33 of ORF1. Studies using antiserum targeted against amino acid residues 6 to 27 indicated that this region (NT) is absent from viral p100 and p190, while pulse-labelling experiments showed that NT is present in nascent p100 synthesized in GLV-infected *Giardia lamblia* but removed subsequently. In contrast, this region was retained in the two viral proteins synthesized in vitro, and it was not removed upon prolonged incubation or inclusion of microsomal fraction in the in vitro translation reaction mixtures. These results suggest that endoplasmic reticulum is not involved in the protein processing and that the precursors of p100 and p190 are incapable of cleaving themselves or each other. This specific cleavage was reproduced when lysates from GLV-infected *G. lamblia* were added, but not those from uninfected cells. The cleavage activity was relatively insensitive to phenylmethylsulfonyl fluoride, but it was inhibitable by leupeptin or E-64, two known specific inhibitors of cysteine protease. The possible origin of this processing activity is discussed.

Many viruses produce proteases for maturational proteolytic cleavage of viral precursor polyproteins. These phenomena have been well documented for the DNA viruses (4), retroviruses (15), and positive-strand RNA viruses (19). Among double-stranded RNA (dsRNA) viruses, the $\sigma 3$ protein of the multisegmented reovirus is known to act as a protease to cleave a myristoylated protein, $\mu 1$, to generate another outer capsid protein, $\mu 1C$ (20). The dsRNAs of bisegmented infectious bursal disease virus and infectious pancreatic necrosis virus each encode a protease used during gene expression (6). In another instance, the hypovirulence-associated dsRNA of the chestnut blight fungal pathogen *Cryphonectria parasitica* encodes an autocatalytic papain-like protease, p29, from one reading frame (3) and an autocatalytic protease, p48, from yet another reading frame (17). However, the presence of a virus-encoded protease may not be universal among the dsRNA viruses. For example, a member of the totivirus family, the dsRNA killer virus (ScV) of *Saccharomyces cerevisiae*, has not been reported to be proteolytically processed and is without a protease gene as determined by sequence analysis (25). These virus-like particles are believed to be descended from the ancestral class I positive-stranded RNA viruses through deletions of the viral RNA helicase genes and the protease genes (10). This family of dsRNA viruses has, typically, an icosahedral capsid and a small nonsegmented dsRNA genome encoding a capsid protein and a *gag-pol*-like fusion protein through a (–1) translational frameshift (23).

One of these dsRNA viruses from the *Totiviridae* family, giardavirus (GLV), infects specifically the trophozoites of an

anaerobic flagellate, *Giardia lamblia*. It is among the smallest and simplest of viruses. The virion is a 36-nm-diameter non-enveloped icosahedron (22) comprising a linear dsRNA of 6,277 bp (24, 26), a major polypeptide of 100 kDa (p100), and a less abundant polypeptide of 190 kDa (p190) (14). Previous studies have shown that the GLV genome has two large open reading frames (ORFs) that overlap by 220 nucleotides (nt) (24). The capsid protein p100 is encoded by ORF1, whereas p190 is synthesized as a *gag-pol*-type frameshift fusion protein of ORFs 1 and 2 with all known motifs characteristic of RNA-dependent RNA polymerases (RDRP) found in the C-terminal domain of ORF2 (24). In all six reading frames throughout the entire viral dsRNA, there is no sequence that bears significant homology to any known protease.

In our present investigation, we noted that the N-terminal 32 amino acid residues derived from ORF1 were missing from both p100 and p190 of mature GLV but that they were part of the p100 precursor protein when it was first synthesized in the infected *G. lamblia*. In vitro translational products synthesized from transcripts of various GLV cDNA constructs in rabbit reticulocyte lysates were also shown to retain the N-terminal peptide, suggesting that p100 and p190 synthesized in vitro are incapable of autoproteolysis. A protease activity that specifically removes the N-terminal peptide for the production of mature GLV proteins was identified and is characterized in this report.

MATERIALS AND METHODS

Cell culture and preparation of cell lysate. *G. lamblia* WB (WB), clone 1, was obtained from R. L. Berens, University of Colorado. It was GLV free and susceptible to GLV infection (13). Culturing conditions have been described previously (22). Unless otherwise specified, *G. lamblia* cells were infected with GLV at a multiplicity of infection of 10 and incubated at 37°C. After 48 h, the cells were washed with phosphate-buffered saline. The lysates of uninfected (WB) cells and GLV-infected (WBI) cells were prepared by sonication in TS

* Corresponding author. Mailing address: Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446. Phone: (415) 476-4085. Fax: (415) 476-3382. Electronic mail address: alwang@cgl.ucsf.edu.

buffer (10 mM Tris-HCl [pH 7.6], 150 mM NaCl), and the protein concentration was determined by the Bio-Rad protein assay.

Preparation of antisera and Western immunoblots. The following peptides were synthesized by using the ABI system at the Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan, conjugated to keyhole limpet hemocyanin (9), and used to immunize rabbits: (i) peptide NT, GYGNSPLNPYGFASLHRG GLN (amino acids [aa] 6 to 27 in ORF1); and (ii) peptide NNT, AGRONVPV LQQNQENEDNHALGGSED (aa 66 to 91 in ORF1). Antiserum against GLV was prepared by purifying GLV (21) and using intact purified virus to immunize rabbits by a standard protocol (9). GLV was isolated by lysing the infected *G. lamblia* WB cells in 1% Nonidet P-40 in TSE buffer (10 mM Tris-HCl [pH 8.0], 100 mM NaCl, 1 mM EDTA) and pelleted twice from the supernatant at 65,000 rpm for 45 min in a Beckman TL100.2 rotor. The final virus suspension was fractionated by sodium dodecyl sulfate–6% polyacrylamide gel electrophoresis (SDS–6% PAGE), blotted to Immobilon-P (Millipore), incubated with rabbit antiserum at a dilution of 1:1,000 (α -NT and α -GLV) or 1:2,000 (α -NNT) for 2 h at room temperature, and washed for 45 min in TST buffer (10 mM Tris-HCl [pH 7.6], 150 mM NaCl, 0.5% Tween 20) at room temperature with at least four changes of buffer. The blot was then incubated in a 1:2,000 dilution of horseradish peroxidase-conjugated anti-rabbit immunoglobulin G (Amersham) for 45 min at room temperature and washed as described above. The immune reaction was visualized with the ECL horseradish peroxidase chemiluminescence system (Amersham).

Construction of recombinant plasmids. The 3.1-kb *Pst*I–*Sst*II DNA fragment containing the GLV ORF1 cDNA and its 367-nt untranslated sequence from the full-length GLV genomic cDNA clone pGEMGLV (24) was inserted into pBlue-script (Stratagene) downstream from the T3 promoter to yield pBS/gag. For ORF1 without the GLV untranslated sequence, pGEMGLV was mutated to create a unique *Nde*I site at nt 367 (p23GLV) and digested with *Sst*II. The linearized DNA was filled in with T4 DNA polymerase to create a blunt end before being cut with *Nde*I at nt 367. The resulting 2.7-kb *Nde*I–*Sst*II cDNA fragment consisting of nt 368 to 3101 of the GLV capsid gene was ligated to the linearized 3.7-kb *Nde*I–*Eco*RV fragment of pCITE-3a(+) (Novagen) to yield the recombinant plasmid pCITE/gag. In this construct, the capsid gene was downstream of a T7 promoter and the encephalomyocarditis virus (EMCV) 5' untranslated region.

To clone the ORF2 cDNA, an *Sst*II–*Nru*I cDNA fragment was inserted into the *Eco*RI and *Eco*RV sites of pCITE-3a(+) after treatment of both DNAs with T4 DNA polymerase. The insert in this construct, pCITE/pol, consists of GLV nt 3101 to 5976 (corresponding to aa 911 to 1870 in the fused ORF1-ORF2). To clone cDNA containing the overlapping ORFs in differing reading frames, p23GLV was cleaved with *Nde*I and *Nru*I, and a DNA fragment containing nt 368 to 5976 was gel purified. The DNA fragment was then inserted into the *Nde*I and *Eco*RV sites of the expression vector pCITE-3a(+) to produce pCITE/gag-pol. pCITE/gag Δ pol is a derivative of pCITE/gag-pol obtained by introducing an extra C nucleotide into the position between 2916 and 2917 via a standard procedure of site-directed mutagenesis (Amersham). The nucleotide insertion fused the two ORFs to express only the gag-pol fusion protein.

In vitro transcription and translation. Recombinant DNA plasmids were linearized at the restriction site 3' to the coding sequences. Runoff transcription reactions with T3 or T7 RNA polymerase were performed as described by the manufacturer (Promega). For capped transcription, 5 μ l of 10 mM m⁷GpppG was added to the reaction mixture and the final concentration of GTP was reduced to 50 μ M. The synthesized RNA was purified by phenol extraction and ethanol precipitation in the presence of 1 M ammonium acetate and then dissolved in diethyl pyrocarbonate-treated H₂O. The purified RNA was analyzed by 1.0% agarose-formaldehyde gel electrophoresis and translated in an RNA-dependent rabbit reticulocyte lysate (Promega) in the presence of 0.8 mCi of [³⁵S]methionine per ml (50 mCi/mmol). The reactions were carried out with 0.5 μ g of RNA transcript in a final volume of 50 μ l at 30°C in the presence or absence of 7.2% (vol/vol) canine pancreatic microsomal membranes (Promega). Translation was terminated by adding DNase-free RNase (Boehringer Mannheim) to 0.01 U/ml and cycloheximide (Sigma) to 6 mM. The translation products were then aliquoted for proteolytic processing and protein analysis by SDS-PAGE and autoradiography.

Pulse-chase analysis of GLV-infected *G. lamblia* cells. *G. lamblia* WB cells were infected with GLV at a multiplicity of infection of 500 as described previously (21). After incubation at 37°C for 16 h, the culture medium was replaced with methionine-free RPMI 1640 (Gibco), and the incubation was continued for another 30 min before the culture was pulsed with 200 μ Ci of [³⁵S]methionine to reach 2 μ Ci/ μ l for 15 min. The reaction mixture was then chased with a 20-fold concentration of unlabeled methionine for 30 min. At various times thereafter, samples of cells were lysed in TNE (10 mM Tris-HCl [pH 8.0], 100 mM NaCl, 1 mM EDTA) with 1% Triton X-100. The lysate was clarified by a 15-min centrifugation at 15,000 $\times$ g and 4°C, and the supernatant was diluted fivefold with radioimmunoprecipitation assay buffer (20 mM Tris-HCl [pH 7.5], 100 mM NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS) and analyzed for the presence of GLV protein by immunoprecipitation with region-specific antisera, SDS-PAGE, and autoradiography as described below.

Proteolytic processing and immunoprecipitation. To assay for possible proteolytic processing of GLV protein, the in vitro-translated, radiolabelled protein was diluted 20-fold in processing buffer (50 mM HEPES [N-2-hydroxyethyl-

FIG. 1. Western blots of GLV protein reacting with immunosera targeted against synthetic peptides. (A) ORF1 and ORF2 of the GLV genome. Dotted box, coding regions for capsid protein; striped box, RNA polymerase domain. The positions of synthetic peptides, C, NT, and NNT, are underlined. (B) Computer-predicted N-terminal amino acid sequence of ORF1. The N-terminal sequence found by Edman degradation in 2 to 5% of the GLV p100 is shown in boldface and underlined. (C) Immunoblots of GLV proteins from purified mature virus reacting with antiserum to peptide C, NT, or NNT. Cross-reacting minor bands around p100 and p190 are occasional contaminating polypeptides not yet characterized.

piperazine-N'-2-ethanesulfonic acid; pH 7.4], 100 mM NaCl, 2 mM dithiothreitol, 1 mM EDTA), samples were incubated at 30°C for 60 min in the absence or presence of protease inhibitors [leupeptin, *trans*-epoxysuccinyl-leucylamido-(4-guanido)butane (E-64), and/or phenylmethylsulfonyl fluoride (PMSF)], and the reactions were terminated by addition of 1% SDS. The incubated mixtures were then diluted 2.5-fold in radioimmunoprecipitation assay buffer containing a cocktail of protease inhibitors (16 μ g of benzamide per ml, 10 μ g of aprotinin per ml, 10 μ g of leupeptin per ml, 10 μ g of pepstatin per ml, 10 μ g of PMSF per ml) and precleared with protein A-agarose (Gibco) for 15 min on ice. The preabsorbed supernatant (500 μ l) was mixed with 2 to 5 μ l of undiluted antiserum and incubated for 60 min on ice before protein A-agarose was added and the incubation was continued for another 60 min. The resultant protein A complex was pelleted and washed three times with the dilution buffer and twice with 20 mM Tris-HCl, pH 7.5, plus 100 mM NaCl. The final pellet was suspended in 30 μ l of SDS sample buffer and heated at 100°C for 3 min. The beads were removed by centrifugation, and the supernatant was analyzed by SDS-PAGE. Gels were dried for 1 h under vacuum at 80°C before being exposed to X-ray film (Kodak X-Omat) at –70°C.

RESULTS

The N terminus predicted from ORF1 is missing from p100 and p190 in mature GLV. Purified GLV particles consist of two polypeptides, one of 100 kDa (p100) and one of 190 kDa (p190), as determined by SDS-PAGE, with p190 present in quantities approximately 2 to 5% of those of p100 (24). Direct peptide sequencing of the two purified proteins indicated that both p100 and p190 are blocked at their N termini. However, 2 to 5% of the capsid protein p100 was found to be unblocked and it yielded an N-terminal sequence, PENITFDT . . . , that coincides with aa 33 to 40 of ORF1 (Fig. 1B). Two peptides, NT (aa 6 to 27) and NNT (aa 66 to 91), were chemically synthesized and used to generate the respective antisera α -NT

and α -NNT in rabbits (Fig. 1B). Two additional rabbit antisera, α -GLV and α -C, from our previous studies (24) were made available as follows: α -GLV was obtained from rabbits immunized with purified GLV and α -C was targeted to aa 505 to 522 in ORF1. Western blots of purified GLV proteins reacting to these antisera are shown in Fig. 1C. Antisera α -C and α -NNT recognized both p100 and p190, as expected. α -NT, however, failed to react with either p100 or p190, suggesting that the N-terminal aa 6 to 27 derived from the sequence of ORF1 are absent in p100 and p190 of mature GLV particles. This could probably be attributed to one of the two following possibilities. (i) The N termini are cleaved from both proteins during maturation of the virus, or (ii) synthesis of GLV proteins in infected *G. lamblia* starts at amino acid residue 33 of ORF1, which is a proline, instead of residue 1, a methionine.

GLV proteins synthesized in vitro do not exhibit protease activity. In order to examine the possibility that the initial N termini of p100 and p190 are cleaved subsequent to their synthesis, we first looked for potential autocatalytic protease activity in *cis* or protease activities in *trans* in the two precursor proteins of p100 and p190 synthesized in vitro. The 3.1-kb GLV cDNA fragment in pBS/gag containing the complete ORF1 plus its 367-nt leader sequence at the 5' end was linearized with *SsrII* and transcribed in vitro with or without a 5' cap structure (Fig. 2A). The 3.1-kb RNA product was verified by agarose gel electrophoresis (data not shown), purified, and subjected to in vitro translation in rabbit reticulocyte lysate. But no detectable synthesis of polypeptide was observed when the reaction products were subjected to SDS-PAGE (Fig. 2B, lane 2). The 367-bp GLV noncoding leader sequence was then replaced with the encephalomyocarditis virus 5' untranslated region in a new construct, pCITE/gag. Results from in vitro transcription and translation of this plasmid indicate a major translational product with an estimated size of about 100 kDa (Fig. 2B, lane 4), which was immunoprecipitable by α -GLV and α -NNT as well as α -NT and was thus designated the precursor p100 (pp100) (Fig. 2C, lanes 1 to 3). Inclusion of canine pancreatic microsomal membranes in the translation mixture did not alter the size of the translational product (Fig. 2B, lane 3) or its reactivity with α -NT and α -GLV (data not shown). Thus, the protein derived from ORF1 did not have detectable autoproteolytic activity for removing the NT peptide. Nor did the rabbit reticulocyte lysate or canine microsomes contain any protease activity capable of digesting the translational product in a specific manner.

The in vitro transcript from plasmid pCITE/gag-pol linearized at the *NotI* site consisted of the encephalomyocarditis virus 5' untranslatable region and the fragment spanning the entire overlapped ORF1 and ORF2. Its translational products included a major protein band of 100 kDa (Fig. 2B, lanes 6 and 7) and a minor band of 190 kDa. Both were immunoprecipitable with α -NT and α -GLV, and the 190-kDa species was designated the precursor p190 (pp190) (Fig. 2C, lanes 5 and 6). The presence of canine pancreatic microsomal membranes did not alter the profile of proteins thus synthesized (Fig. 2B, lane 7). This finding suggests that the NT peptide was not cleaved from pp100 by pp190 and vice versa. In an effort to produce enough precursor 190-kDa gag-pol protein to further verify this conclusion, plasmid pCITE/gag Δ pol, which contains an additional C between nt 2916 and 2917 to fuse ORFs 1 and 2 into a single ORF, was transcribed and translated in vitro. The 190-kDa protein, a major product from the in vitro translation (Fig. 2B, lanes 8 and 9), was immunoprecipitable by α -GLV, α -NNT, and α -NT (Fig. 2C, lanes 7 to 9), suggesting that the fusion protein was pp190, which lacks the autocatalytic protease activity to cleave its own N-terminal peptide. An addi-

FIG. 2. Autoradiograms of SDS-PAGE analysis of the in vitro translation products. (A) Schematic representation of the GLV genome, predicted protein-coding regions, and various DNA constructs used in this study. The GLV genome is shown as a thick line. Thin lines, GLV untranslated regions; open boxes, ORFs. Constructs expressing specific regions of the GLV genome are shown below the diagram. Thick lines, encephalomyocarditis virus 5' untranslated region. (B) In vitro translation products labelled with [35 S]methionine in rabbit reticulocyte lysate. Transcripts from plasmid constructs were translated in the absence (lanes 1, 2, 4, 6, and 9) or presence (lanes 3, 7, and 8) of canine microsomal membranes and analyzed by SDS-PAGE-autoradiography. (C) Immunoprecipitation of GLV polypeptides translated in vitro. Products were immunoprecipitated with antisera against GLV (α -GLV) or the peptide NT (α -NT) or NNT (α -NNT) and analyzed by SDS-PAGE-autoradiography. pp190 and pp100 are precursors of the 190- and 100-kDa proteins, respectively; pol is the translation product of pCITE/pol. Positions of molecular mass standards (Gibco BRL) in kilodaltons are indicated.

tional construct, pCITE/pol, which has the entire ORF2 cDNA that is presumably the RDRP domain of the fusion protein p190, yielded a 107-kDa protein (pol) (Fig. 2B, lane 5) immunoprecipitable only by α -GLV, as expected (Fig. 2C, lane 4). When the in vitro transcripts from pCITE/pol and pCITE/gag were cotranslated in vitro, two proteins of 107 and 100 kDa were synthesized (Fig. 3A, lane 3). As expected, only the 100-kDa band was recognized by either α -NT or α -NNT (Fig. 3A, lanes 1 and 2). By a similar approach, cotranslation of the transcripts from pCITE/gag Δ pol and pCITE/gag resulted in two major proteins of 190 and 100 kDa (Fig. 3B, lanes 1 to 3). In both cases, immunoprecipitation experiments indicated that the 100- and 190-kDa proteins can both be brought down by α -NT, identifying them as pp100 and pp190, respectively, and suggesting the absence of *trans*-acting proteolytic digestion among the two proteins and the RDRP portion of p190 (Fig.

FIG. 3. SDS-PAGE analysis of immunoprecipitated GLV proteins either synthesized separately in vitro or cotranslated with mixed RNA transcripts. (A) Immunoprecipitation of cotranslated products from pCITE/gag and pCITE/pol RNAs, both labelled with [35 S]methionine (lanes 1 to 3), or 35 S-labelled products from pCITE/gag RNA incubated with unlabelled products from pCITE/pol RNA (lanes 4 to 6). (B) Immunoprecipitation of 35 S-labelled products from mixed transcripts of pCITE/gag and pCITE/gag Δ pol (lanes 1 to 3) or 35 S-labelled pCITE/gag products incubated with unlabelled pCITE/gag Δ pol products (lanes 4 to 6). pp190 and pp100 are precursor 190- and 100-kDa proteins, respectively; pol is the translation product of pCITE/pol.

3). When the three proteins were synthesized separately in vitro and only pp100 was radiolabelled, the radiolabelled pp100 remained immunoprecipitable by α -NT after being incubated with unlabelled pol or pp190 at 30°C for 60 min (Fig. 3A and B, lanes 4). These results again suggest that the N-terminal peptide in pp100 was intact even after prolonged incubation with pp190 or the RDRP domain of p190. We therefore conclude that none of the three GLV proteins synthesized in vitro possessed any *cis*- or *trans*-acting proteolytic function.

Synthesis of GLV protein in infected *G. lamblia* is initiated at the first methionine. In order to examine the actual start site of GLV protein synthesis in *G. lamblia*, we pulse-labelled the proteins in GLV-infected cells with [35 S]methionine and immunoprecipitated the newly synthesized GLV proteins with region-specific antisera. The result with α -GLV indicated that a 100-kDa protein was predominantly labelled after the 15-min pulse (Fig. 4). This protein could be immunoprecipitated by α -NT, suggesting that the sequence encoding peptide NT, located at the very beginning of ORF1, was expressed in the GLV-infected cells and that the viral protein synthesis started most likely at the first methionine of ORF1. Following an additional 30-min chase with unlabelled methionine, two major protein bands in the 100-kDa region were immunoprecipitated by α -GLV, whereas only the upper protein band could be immunoprecipitated by α -NT (Fig. 4, lanes 3 and 4). These results, together with data presented in Fig. 1 and 2, suggest clearly that the NT peptide derived from ORF1 was translated as the N terminus of the precursor p100, and probably also the precursor p190 as well, in the GLV-infected *G. lamblia* cells and that it was subsequently cleaved off in mature GLV particles. We examined GLV-infected cell lysate in high-percentage SDS-PAGE followed by Western blotting to look for this cleaved small fragment but did not find it. This fragment either was lost because of its small size (32 aa) or was rapidly degraded in the lysate, as many cleaved fragments tend to be.

Identification of the protease activity in GLV-infected *G. lamblia* cells responsible for the cleavage of viral protein. To

FIG. 4. Pulse-chase analysis of viral proteins in GLV-infected *G. lamblia*. *G. lamblia* cells were infected with GLV at a multiplicity of infection of 500. After 16 h, the cells were pulse-labelled with [35 S]methionine for 15 min, and GLV-specific proteins were isolated from the cells by immunoprecipitation with α -NT or α -GLV. After a chase with unlabelled methionine for 30 min, the cell lysate was again precipitated with the same antisera. pp100 is the 100-kDa precursor protein.

search for the protease responsible for cleaving the GLV precursor protein in GLV-infected cells, we developed an assay for the proteolytic processing of pp100. Radiolabelled pp100 was produced by in vitro translation of 0.5 μ g of the pCITE/gag transcript in a final volume of 50 μ l. Ten microliters of the radiolabelled pp100 was mixed with 0.8 μ g of total protein from either uninfected (WB) or GLV-infected (WBI) *G. lamblia* cell lysate, and the mixtures were diluted to 200 μ l with processing buffer. After incubation for 60 min at 30°C, the mixtures were immunoprecipitated with α -NT, α -NNT, or α -GLV and analyzed by SDS-PAGE and autoradiography (Fig. 5). pp100 remained unchanged when it was incubated with only Tris-saline (Fig. 5, lanes 1 to 3), but it was completely degraded by WB (lanes 16 to 18) or WBI (lanes 4 to 6) lysate in the absence of protease inhibitors. This degradation was effectively blocked by either leupeptin at 1 μ M (lanes 10 to 12 and 22 to 24) or E-64 at 0.8 μ g/ml (Fig. 5, lanes 13 to 15). Interestingly, PMSF at 0.5 mM largely inhibited nonspecific degradations of pp100 by the WB lysate, but it allowed the WBI lysate to generate a smaller protein band which could be immunoprecipitated by α -NNT and α -GLV (Fig. 5, lanes 8 and 9) but not by α -NT (lane 7) and which was thus most likely the mature p100. In comparing PMSF plus WB lysate (lanes 19 to 21) with PMSF plus WBI lysate (lanes 7 to 9), one sees that pp100 remained unchanged with the WB lysate but that specific cleavage of pp100 to form p100 was apparent with the WBI lysate. Thus, the specific cleavage activity converting pp100 to p100 is resistant to PMSF and present only in WBI lysate.

The levels of pp100 and p100 were examined more closely in subsequent experiments. It was found that p100 increased in intensity when the incubation time of pp100 with WBI lysate was prolonged from 15 min to 120 min in the presence of 0.5 mM PMSF, while the level of pp100 decreased concomitantly (Fig. 6). These experiments were repeated several times with WBI lysate generated at different times after GLV infection and with different amounts of lysate to construct a reproducible time course (Fig. 6) and stoichiometry (data not shown). In contrast, no change in the level of pp100 incubated with the WB lysate in the presence of PMSF was observed (lanes 19 to 21). This observation demonstrated that a specific protease activity, not found in rabbit reticulocyte lysate, canine microsomal membrane fraction, or uninfected *Giardia* cells, is responsible for the specific proteolytic modification of the GLV proteins. Since both leupeptin and E-64 are specific inhibitors of cysteine proteases (8), this processing activity probably comes from the cysteine protease family. Apparently, this enzyme activity can function in *trans* to cleave a substrate that is synthesized separately. It was also found to be stable in crude *Giardia* cell lysates upon freezing.

FIG. 5. Proteolytic processing of GLV capsid protein synthesized in vitro by *G. lamblia* cell lysate. RNA transcript from pCITE/gag (Fig. 2) was translated in rabbit reticulocyte lysate in the presence of [³⁵S]methionine and then incubated with lysates of virus-free *G. lamblia* (WB) or GLV-infected *G. lamblia* (WBI) cells in the presence or absence of protease inhibitors, after which the reaction mixtures were subjected to immunoprecipitation with rabbit antisera. a, α -NT; b, α -NNT; c, α -GLV. Protease inhibitors leupeptin, *trans*-epoxysuccinyl-leucylamido-(4-guanidino)butane (E-64), and PMSF were added as indicated. pp100 is the 100-kDa precursor protein.

DISCUSSION

We have analyzed the in vitro translational products of transcripts derived from the GLV cDNA. The results demonstrated that the 367 nt in the 5' untranslated regions of the viral transcripts are unsuitable for use as the translational initiation site in rabbit reticulocyte lysate, even though they must function well in GLV-infected *G. lamblia* cells. This discrepancy may reflect the fact that *G. lamblia* is among the most ancient eukaryotic microorganisms and that it differs in many respects from mammalian cells (18). Its small subunit rRNA bears a structure significantly different from that of mammalian ribosomes (2). The mRNAs observed in *G. lamblia* often carry in their 5' untranslated regions only 1 to 6 nt (1), which are apparently sufficient for triggering the process of translation.

Although most of the mature GLV protein is blocked at the N terminus, we have sufficient evidence to indicate that the putative capsid protein p100 is initially synthesized from the first methionine residue in ORF1 and that the N-terminal 32 aa are removed subsequently. This process, presumably essential for viral assembly and maturation, is catalyzed by a cysteine protease-like activity not found in *G. lamblia* cells prior to GLV infection. Nor is it detectable among the proteins encoded by the GLV dsRNA genome and synthesized in vitro.

The lack of demonstrable self-cleavage activity by pp100 or pp190 can be attributed to the following two very different possibilities. (i) The mature GLV proteins synthesized in vivo are self-cleaved, but the precursors synthesized in vitro lack proteolytic activity because of improper folding or anomalous modification in a heterologous system, or (ii) the cleavage activity comes from a *Giardia* enzyme whose expression or

activity is upregulated after GLV infection. It is interesting to note that the fragment of cDNA encoding pp100, i.e., GLV nt 368 to 3110, has been cloned into baculovirus vector and expressed in Sf9 insect cells (16). The recombinant protein thus synthesized had a molecular size of 100 kDa and was recognized by both α -GLV and α -NT, suggesting that, similar to the case with the product from in vitro translation, the NT region was also retained in the protein. The results from these two heterologous protein expression systems suggest that only GLV-infected *G. lamblia* is capable of providing the necessary elements for GLV maturation.

There are many examples of host factors playing important roles in virus replication. Most notably, cellular proteases, in addition to viral ones, are required for maturation cleavage of many myxoviruses and coronaviruses (reviewed in reference 11). Besides the individual proteases, host metabolic or degradative pathways can also be involved. For example, the E₁ and E₂ glycoproteins of togaviruses must be processed by host signal proteases before myristoylation and glycosylation of the viral proteins can occur (12). In the case of Sindbis virus, the viral polymerase must be degraded through the cellular ubiquitin pathway to stop the synthesis of viral RNA and for the virus to proceed to the next stage of the replication cycle (5). However, the expression of all of these host proteases or other metabolic pathways is already present before the virus infection. The degradation of Sindbis virus polymerase can be observed even in the rabbit reticulocyte lysate. In contrast, the GLV maturation cleavage cannot take place in uninfected *G. lamblia* cell lysates. Our present data are consistent with the hypothesis that this cleavage activity is attributable to a GLV-induced expression or activation of a cellular protease in *G. lamblia*. Alternatively, the precursor GLV polypeptides may require a specific cellular chaperonin similar to the GroEL of *Escherichia coli* (7) that is present in *G. lamblia* but missing from rabbit reticulocyte lysate or Sf9 cells for proper folding of the viral precursor protein to attain proteolytic activity. Elucidation of this interesting possibility must await analysis of the gene expression in *G. lamblia* cells after GLV infection and closer examination of the intracellular events that regulate GLV replication and protein biogenesis.

In summary, the results presented in this report provide a description of the novel processing event required for the biosynthesis of the viral capsid protein. Further studies will be needed to identify and isolate the responsible protease and to define the biological role of the N-terminal modification of capsid protein in viral encapsidation. As of now, this is also the first example from among the nonsegmented dsRNA totiviruses in which the virus relies entirely on a host factor for its proteolytic maturation. For the time being, however, there is

FIG. 6. Processing time course of the 100-kDa precursor protein in GLV-infected *G. lamblia* (WBI) lysate. The 100-kDa radiolabelled precursor protein was generated in vitro and incubated with WBI lysate in the presence of 0.5 mM PMSF. Samples were taken at various times as indicated and analyzed by SDS-PAGE after immunoprecipitation with α -NT (lanes 1, 3, 5, 7, 9, and 11) or α -GLV (lanes 2, 4, 6, 8, 10, and 12). pp100 is the 100-kDa precursor protein.

no indication from a database search that the GLV genome could encode any protease-like protein among all possible ORFs.

ACKNOWLEDGMENTS

We thank Hung-Min Yang for the construction of pGEMGLV and p23GLV, Juan Wang for the gagΔpol insertion mutant, and Jürg M. Sommer and Shaobing Hua for helpful advice and discussion.

This work was supported by grant AI-30475 from the National Institutes of Health.

REFERENCES

- Adam, R. D. 1991. The biology of *Giardia* spp. Microbiol. Rev. 55:706–732.
- Boothroyd, J. C., A. Wang, D. A. Campbell, and C. C. Wang. 1987. An unusually compact ribosomal DNA repeat in the protozoan *Giardia lamblia*. Nucleic Acids Res. 15:4065–4084.
- Choi, G. H., D. M. Pawlyk, B. Rae, R. Shapira, and D. L. Nuss. 1993. Molecular analysis and overexpression of the gene encoding endothiopsin, an aspartic protease from *Cryptosporidium parvum*. Gene 125:135–141.
- Covey, S. N. 1991. Pathogenesis of a plant pararetrovirus: CaMV. Semin. Virol. 2:151–159.
- de Groot, R. J., T. Rumenapf, R. J. Kuhm, E. G. Strauss, and J. H. Strauss. 1991. Sindbis virus RNA polymerase is degraded by the N-end rule pathway. Proc. Natl. Acad. Sci. USA 88:8967–8971.
- Dougherty, W. G., and B. L. Semler. 1993. Expression of virus-encoded proteinase: functional and structural similarities with cellular enzymes. Microbiol. Rev. 57:781–822.
- Georgopoulos, C., R. W. Hendrix, S. R. Casjens, and A. D. Kaiser. 1973. Host participation in bacteriophage lambda head assembly. J. Mol. Biol. 76:45–60.
- Hanada, K., M. Tamai, T. Adachi, K. Oguma, K. Kashiwagi, S. Ohmura, E. Kominami, T. Towatari, and N. Katunuma. 1983. Characterization of the three new analogs of E-64 and their therapeutic application, p. 25–36. In N. Katunuma (ed.), Proteinase inhibitors: medical and biological aspects. Japan Sci. Soc. Press, Tokyo.
- Harlow, E., and D. Lane. 1988. Antibodies: a laboratory manual. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
- Koonin, E. V., and V. V. Dolja. 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. Crit. Rev. Biochem. Mol. Biol. 28:375–430.
- Krausslich, H.-G., and E. Wimmer. 1988. Viral proteases. Annu. Rev. Biochem. 57:701–754.
- McDonald, H., T. C. Hobman, and S. Gillam. 1991. The influence of capsid protein cleavage on the processing of E2 and E1 glycoproteins of rubella virus. Virology 183:52–60.
- Miller, R. L., A. L. Wang, and C. C. Wang. 1988. Identification of *Giardia lamblia* isolates susceptible and resistant to infection by the double-stranded RNA virus. Exp. Parasitol. 66:118–123.
- Miller, R. L., A. L. Wang, and C. C. Wang. 1988. Purification and characterization of the *Giardia lamblia* double-stranded RNA virus. Mol. Biochem. Parasitol. 28:189–196.
- Oroszlan, S., and R. B. Luftig. 1990. Retroviral proteinase. Curr. Top. Microbiol. Immunol. 157:153–185.
- Sepp, T., A. L. Wang, and C. C. Wang. Expression of the giardavirus putative capsid gene in Baculovirus system. Exp. Parasitol., in press.
- Shapira, R., and D. L. Nuss. 1991. Gene expression by a hypovirulence-associated virus of the chestnut blight fungus involves two papain-like protease activities. Essential residues and cleavage site requirements for p48 autoproteolysis. J. Biol. Chem. 266:19419–19425.
- Sogin, M. L., J. H. Gunderson, H. J. Elwood, H. J. Alonso, and H. J. Peattie. 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. Science 243:75–77.
- Strauss, E. G., and J. H. Strauss. 1991. RNA viruses: genome structure and evolution. Curr. Opin. Genet. Dev. 1:485–493.
- Tillotson, L., and A. J. Shatkin. 1992. Reovirus polypeptide σ3 and N-terminal myristoylation of polypeptide μ1 are required for site-specific cleavage to μ1C in transfected cells. J. Virol. 66:2180–2186.
- Wang, A. L., R. L. Miller, and C. C. Wang. 1988. Antibodies to the *Giardia lamblia* double-stranded RNA virus major protein can block the viral infection. Mol. Biochem. Parasitol. 30:225–232.
- Wang, A. L., and C. C. Wang. 1986. Discovery of a specific double-stranded RNA virus in *Giardia lamblia*. Mol. Biochem. Parasitol. 21:269–276.
- Wang, A. L., and C. C. Wang. 1991. Viruses of the protozoa. Annu. Rev. Microbiol. 45:251–263.
- Wang, A. L., H. M. Yang, K. A. Shen, and C. C. Wang. 1993. Giardavirus double-stranded RNA genome encodes a capsid polypeptide and a gag-pol-like fusion protein by a translation frameshift. Proc. Natl. Acad. Sci. USA 90:8595–8599.
- Wickner, R. B. 1992. Double-stranded and single-stranded RNA viruses of *Saccharomyces cerevisiae*. Annu. Rev. Microbiol. 46:347–375.
- Wu, C.-H., C. C. Wang, H. M. Yang, and A. L. Wang. Nucleotide sequences at all four termini of the giardavirus dsRNA. Gene, in press.