Processing of Rotavirus Glycoprotein VP7: Implications for the Retention of the Protein in the Endoplasmic Reticulum

S. Clare Stirzaker, Peter L. Whitfeld, David L. Christie,* A. Richard Bellamy,‡ and Gerald W. Both

Division of Molecular Biology, Commonwealth Scientific and Industrial Research Organisation (CSIRO), North Ryde, New South Wales 2113, Australia; and Departments of ‡Cell Biology and *Biochemistry, University of Auckland, Auckland, New Zealand

Abstract. Rotaviruses are icosahedral particles that assemble in the lumen of the endoplasmic reticulum (ER). The viral glycoprotein, VP7, is also directed into this compartment and is retained for assembly onto the surface of viral cores. VP7 is therefore a resident ER glycoprotein with a luminal orientation.

The VP7 gene possesses two potential in-frame initiation codons, each preceding a hydrophobic domain. Mature VP7 is derived from a precursor by cleavage but the site of cleavage has not been determined because viral VP7 has a blocked amino terminus. Using site-directed mutagenesis of the gene and in vitro transcription and translation systems, we have investigated the synthesis and processing of the primary products synthesized from each initiation codon. Proteins translated from either codon were processed in vitro to yield products indistinguishable in size. The primary translation products therefore appeared to be cleaved at the same site. The site was located empirically between Ala50 and Gln51 and mutation of the gene to convert Ala50→Val prevented processing. Amino-terminal sequence analyses of proteins synthesized in vitro, and characterization of an amino-terminal fragment of VP7 purified from virus unequivocally established Gln51 as the amino-terminal residue. Pyroglutamic acid was tentatively identified as the blocking group.

Processing of VP7 therefore removes both amino-terminal hydrophobic domains from the protein. Some other mechanism not requiring the presence of these hydrophobic sequences must account for the retention of this novel glycoprotein in the ER.

ROTAVIRUSES are icosahedral viruses possessing a distinctive morphogenetic pathway (11). The inner core of the virus assembles in the cytoplasm in viroplasmic inclusions and buds through the membrane of the endoplasmic reticulum (ER),[1] becoming enveloped in the process. The membrane envelope is later lost by an unknown mechanism and the outer proteins are assembled on the surface of the core (7). Glycoprotein VP7, the major serotype antigen of rotaviruses, is directed to the lumen of the ER where it is retained (20), apparently as an integral membrane protein (10). The incomplete virus particles therefore probably acquire VP7 either during or after the budding process.

The VP7 gene has now been cloned and sequenced for five serotypes of animal and human rotaviruses (8, 9). In each case the nucleotide sequence predicts an open reading frame of 326 amino acids beginning with an initiation codon with a weak consensus sequence (12, 13). A second, in-frame initiation codon with a strong consensus sequence lies 30 codons downstream. Each initiation codon precedes a region of hydrophobic amino acids which is able to direct VP7 to the ER (27). During translocation VP7 is cleaved (6), but the exact cleavage site was not identified because the viral protein had a blocked amino terminus (1, 15). Depending on where VP7 translation begins and where the protein is cleaved during translocation, the gene could potentially produce two proteins, one perhaps in minor amount, which might have subtly different functions. In fact, two species of VP7 have been reported to exist in virus particles (4). Multiple in-frame initiation codons that generate distinct polypeptides have also been observed in other viruses (12).

Most proteins destined for export from the cell are translocated across the ER membrane with concomitant cleavage of the amino-terminal signal sequence (26). In some cases this does not occur, resulting in the formation of a class II membrane protein which is retained in the membrane via a combined signal/anchor domain (28). VP7 would thus be regarded as a class II protein if processing preserved the second hydrophobic domain and it served as a membrane anchor. However, because the virus matures in the lumen of the ER and in its final form is nonenveloped, it might be expected that VP7 could be processed to remove both hydrophobic domains before assembly on the surface of the virus.

A series of empirical rules to predict cleavage sites in membrane proteins has been proposed (24, 25). When these are applied to VP7 (Fig. 1) a cleavage site after residue 50

Address reprint requests to Dr. Both, CSIRO Division of Molecular Biology, P.O. Box 184, North Ryde, NSW 2113, Australia.

1. *Abbreviations used in this paper*: endo H, endoglycosidase H; ER, endoplasmic reticulum.

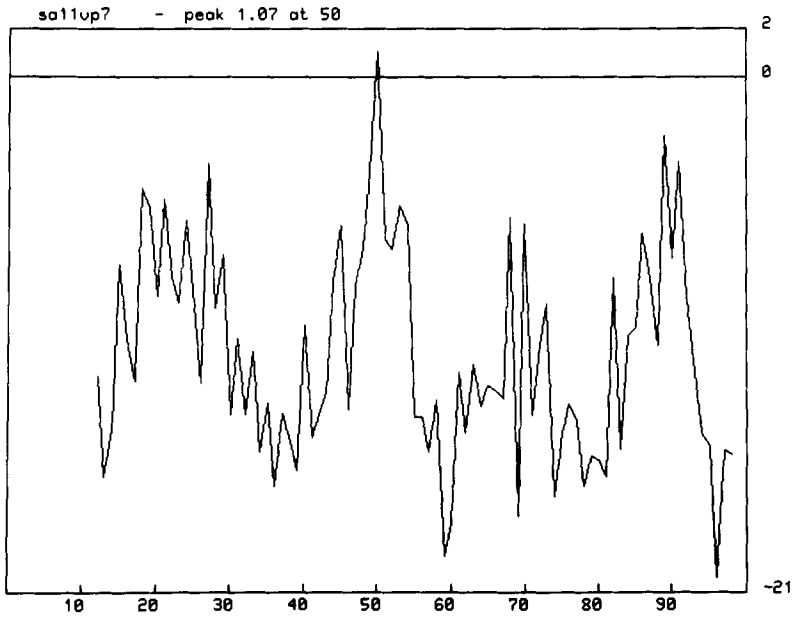

Figure 1. Location of possible signal peptidase cleavage sites within the amino-terminal 100 residues of VP7 using empirical rules formulated from the study of known cleavage sites by von Heijne (25) and a computer program written by C. Bucholtz. The vertical axis gives the score calculated by summing the natural logarithms of the weight-matrices for each residue (25). The horizontal axis shows amino acid numbers in the open reading frame.

is identified. This predicts that mature VP7 would possess amino-terminal glutamine and lack both hydrophobic domains. The availability of a cloned copy of the VP7 gene (2) and the relative ease of constructing mutations has led us to investigate the use of the cleavage site by site-directed mutagenesis. This approach, coupled with partial amino terminal sequence analysis of the amino-terminal region of VP7 has enabled us to unequivocally identify the cleavage site as residue 50.

Materials and Methods

Mutation and Manipulation of VP7 Genes

A full-length cloned copy of the VP7 gene (i.e., segment 9 of SA11) and mutants of the VP7 gene, 1-14 and 27S (Fig. 2), were inserted into the SV40 expression vector pJC119 (20, 27). Insertion of the VP7 gene and its deletion mutants into the Bluescribe vector (Stratagene, San Diego, CA), and the construction of further point mutations Z, S, and T (Fig. 2, Table I) has been described (20, 27).

Additional mutations were introduced into the VP7 gene as follows:

(*a*) The M mutation, which converted the second AUG codon to UUG (Fig. 2, Table I), was introduced by a strategy similar to that described for the T mutation (27). Briefly, the 5'-terminal BamHI fragment from the Z variant of the gene was subcloned into M13mp11 and single-stranded template DNA was prepared. An oligonucleotide (19-mer) corresponding to the coding strand of the VP7 gene was synthesized (DNA synthesizer, model 380A, Applied Biosystems, Inc., Foster City, CA) with a single-base mismatch at base 136 (Table I). The oligonucleotide was elongated by DNA polymerase (Klenow) in the presence of T4 DNA ligase using the single-stranded M13 DNA as template (17). Mutant colonies were identified by differential hybridization of the radiolabeled oligonucleotide at a temperature near its Tm and the presence of the mutation was confirmed by nucleotide sequencing. Double-stranded DNA was prepared, the 5' XhoI-NcoI ~230-bp fragment purified and the modified sequences incorporated into the VP7 gene in the SV40 vector pJC119 by three-fragment ligation (20, 27).

(*b*) Mutation C3 (Fig. 2, Table I), designed to inhibit potential cleavage by signal peptidase, was a point mutation in VP7 changing codon GCA (Ala50) to GTA (Val). The mutation was introduced independently into both the full-length VP7 gene 9ZMT, and the deletion mutant 1-14T (Fig. 2). Single-stranded DNA from these genes in the Bluescribe M13+ vector was prepared using protocols recommended by the supplier and a 19-base oligonucleotide of plus-stranded sense, with a mismatch at position 197, was used to introduce the mutation.

Transfection of DNA and Detection of Expressed Proteins

Transfection of plasmid DNAs into cone outer segment (COS) cells, recov-

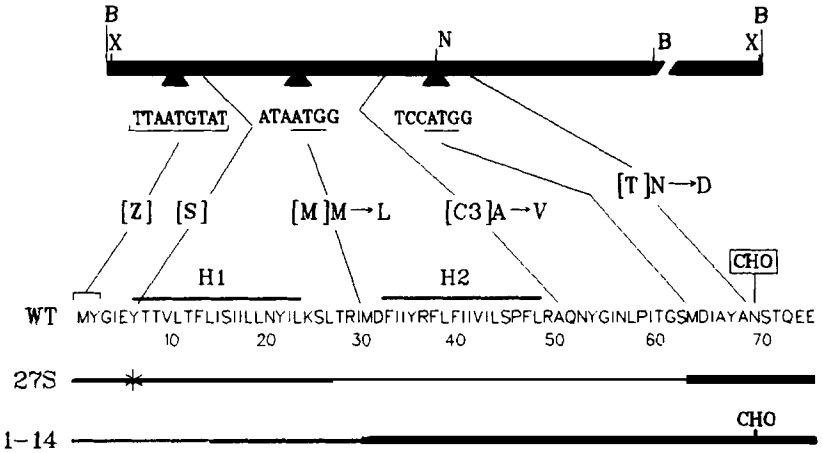

Figure 2. Partial maps for wild-type and variant VP7 genes and proteins. The top of the figure shows the position of the first three in-frame initiation codons and the location of mutations Z, S, M, C3, and T relative to key restriction endonuclease sites within the gene (*B*, BamHI; *N*, NcoI; *X*, XhoI). The amino-terminal 74 residues of wild-type (*WT*) VP7 and mutations in them are shown using the single-letter amino acid code. Hydrophobic regions H1 and H2 are overlined. For the deletion mutants 27S and 1-14, heavy lines indicate amino acid sequence translated from the gene. Former coding regions not translated because of an upstream terminator (*, 27S) or the deletion of an initiation codon (1-14) are shown as thin lines. Regions of DNA deleted are shown as the thinnest line. *CHO*, carbohydrate attached at Asn69.

Table I. Summary of Mutations in the SA11 VP7 Gene

Mutation*	Codon(s) altered	Effect
Z	TTAATGTAT > ATAATGGAT	"Weak" to "strong" first initiation codon plus Tyr2 changed to Asp‡
S	TAC > TAA/G	Tyr6 to STOP‡
M	ATG > TTG	Second initiation codon (Met30) to Leu§
C3	GCA > GTA	Ala50 changed to Val to remove potential signal peptidase cleavage site§
T	AAT > GAT	Asn69 → Asp, removes the only N-linked glycosylation site in SA11 VP7‡

* The wild-type VP7 protein of SA11 virus is encoded in genomic segment 9. Multiple mutations in this gene have been designated in the text by the use of multiple letters (e.g., 9ZMT, Fig. 2, contains three separate mutations in the gene 9 nucleotide sequence).
‡ Whitfeld et al. (27).
§ This work.

ery of expressed VP7 proteins by immunoprecipitation, and their analysis by polyacrylamide gel electrophoresis and autoradiography were carried out as previously described (20, 27).

In Vitro Transcription and Translation

The wild-type VP7 gene and the derived mutant genes were excised from pJC119 with XhoI and subcloned into the SalI site of the Bluescribe M13+ vector under the control of the T7 transcriptional promoter. RNA transcripts were synthesized for 1 h at 37°C, as directed by the supplier, except that m7G(5')ppp(5') G (Pharmacia Fine Chemicals, Uppsala, Sweden) (500 μM) was included in the reaction and the GTP concentration was reduced to 10 μM. The resulting mRNAs were translated using rabbit reticulocyte lysates (Promega Biotec, Madison, WI) in the presence of [^{35}S]methionine (Amersham [Aust.], North Ryde, NSW) and canine pancreatic microsomes (Amersham [Aust.]) when appropriate (27). Conditions for digesting translation products with endoglycosidase H (endo H) or trypsin and the analysis of proteins by gel electrophoresis have been described (4, 14).

Growth of SA11 Rotavirus

SA11 rotavirus was propagated and radiolabeled with [^{35}S]methionine or [^{3}H]mannose in MA104 cells and purified by equilibrium centrifugation on a CsCl gradient (23).

Amino-terminal Sequence Analysis of Proteins

Polypeptides translated and radiolabeled in vitro or [^{35}S]methionine-labeled viral VP7, were separated by electrophoresis on polyacrylamide gels and purified by electroelution (22). The eluted protein was transferred to a siliconized Eppendorf tube and precipitated with 9 vol of methanol overnight at −20°C. Acetylated BSA (2 μg/ml) was added as protein carrier. The precipitated protein was resuspended in 0.1% SDS (50 μl) and submitted to Edman degradation using a microsequencer (model 470A, Applied Biosystems, Inc.).

To remove putative pyroglutamic acid from the amino terminus, eluted proteins were resuspended in deblocking buffer (50 μl) (19), pyroglutamate aminopeptidase (5 U, Sigma Chemical Co., St. Louis, MO) was added, and the sample was incubated at room temperature for 2 h. A further 5 U of enzyme was then added and incubation was continued for a further 2 h. The protein was precipitated and resuspended in 0.1% SDS (50 μl) for Edman degradation. Radioactivity released during each cycle was determined by liquid scintillation counting.

Results

Identification of Proteins Translated from each Initiation Codon

The putative product of the first initiation codon is very poorly translated in rabbit reticulocyte lysates (2, 4), presumably because of the presence of a weak consensus sequence surrounding the codon (13). To enable the full-length protein to be more easily characterized, and to improve expression of VP7 from the first ATG site, the sequence of the gene was modified from 5'..TTAATGT..3' to ..ATAATGG.. (Z mutation, Fig. 2, Table I) (27). Translation from the first AUG codon increased an estimated three- to fivefold in reticulocyte lysates (data not shown).

The identity of VP7 proteins translated from the two potential initiation codons in the gene and a third in-frame codon at position 63 (Fig. 2) was determined by translation of RNA transcripts derived from the full-length gene 9 and the two deletion mutants, 1-14 and 27S. With these transcripts protein synthesis should initiate at the first, second, and third AUG codons, respectively (Fig. 2). The T mutation (Fig. 2, Table I), which removes the glycosylation site, was also incorporated into some genes so that the size of the processed polypeptides could later be examined in the absence of any size change due to glycosylation. The full-length gene 9ZT yielded four proteins in the size range expected for VP7 (Fig. 3a, lane *1*). The smallest of these was usually faint. The bands were all VP7-related in that none of them were seen when no RNA was added to the lysate and all were immunoprecipitated with antiserum directed against VP7 (data not shown). Only the largest product was specific to the first initiation codon because the middle two proteins were synthesized in response to 1-14T RNA (lane *5*) and the smallest was made in the presence of 27S RNA (lane *7*). Inasmuch as both 1-14T proteins disappeared when the M mutation was incorporated in gene 9ZMT (lane *3*), clearly they were translated from the second initiation codon. The reason for the synthesis of two protein products from this codon is not clear. However, the partial amino-terminal sequence of each protein deduced by Edman degradation of [^{3}H]isoleucine-labeled proteins was identical, beginning as expected at Met30 (data not shown). Further, transcripts derived from a 1-14 gene cut at an AccI site 79 bases from the stop codon, encoded two similar, but shorter proteins (data not shown), suggesting that premature termination of translation does not account for their synthesis. In summary, transcripts from various mutated VP7 genes were translated in reticulocyte lysates to identify proteins specific for the first, second, and third in-frame initiation codons.

Cleavage of Proteins Derived from Each Initiation Codon

To determine whether proteins translated from each AUG

codon could be transported across membranes and processed, canine pancreatic microsomes were added to the translation reactions. Entry of proteins into the microsomes was monitored by their resistance to digestion with exogenous trypsin. The size of mutant protein 27S (translated from the third AUG) was unaffected by the addition of microsomes (Fig. 3, *top*, lanes *7* and *8*), and the protein was sensitive to trypsin (27), indicating that it lacked the ability to cross the microsomal membrane. In contrast, the 9ZMT and 1-14T proteins (Fig. 3, *top*, lanes *3* and *5*) translated from the first and second AUG codons, respectively, were processed in the presence of microsomes to a lower molecular mass product of the same apparent size (Fig. 3, *top*, lanes *4* and *6*). This result was confirmed using gene 9ZT, which produced proteins from each AUG; only one new protein appeared when microsomes were added (Fig. 3, *top*, lane *2*). These microsome-processed products were also tested for their resistance to trypsin digestion. In the absence of microsomes, the 9ZMT and 1-14T proteins were trypsin-sensitive (Fig. 3, *bottom*, lanes *3* and *7*). However, the same proteins produced in the presence of microsomes were resistant to digestion with trypsin (Fig. 3, *bottom*, lanes *4* and *8*) indicating that they had crossed the microsomal membrane.

Similar experiments were performed to examine the processing of proteins 9Z, 9ZM, and 1-14 that lacked the T mutation and therefore could be glycosylated. Despite the variation in the size of the precursor proteins translated in the absence of microsomes (Fig. 4, *left*, lanes *3*, *9*, and *12*), the glycosylated proteins produced in the presence of microsomes were identical in size (lanes *4*, *8*, and *11*) and appeared to comigrate with glycosylated VP7 immunoprecipitated from SA11-infected MA104 cells (lane *2*). After digestion with endo H, the resulting unglycosylated polypeptides were also identical in size (lanes *5*, *7*, and *10*) and comigrated with the major species of VP7 present in endo H–treated proteins immunoprecipitated from SA11-infected cells (lane *6*). The three mutant proteins and VP7 from purified virus were also compared after digestion with endo H. Virus yielded a major band of similar size to the processed forms of the three in vitro products, but two minor viral bands were also present (Fig. 4, *right*, lanes *3*, *4*, and *5* vs. *7*). The small difference in migration of the major viral protein relative to the mutant proteins (lanes *3–5*) suggests that it may differ in some subtle way from the major protein in virus-infected cells which appeared to comigrate with the proteins produced in vitro (Fig. 4, *left*, lane *6* vs. lanes *5*, *7*, and *10*). Because microsome-dependent cleavage of the VP7 proteins initiated at the first and second AUG codons appears to produce a single protein, cleavage presumably occurs at a site common to both proteins, i.e., downstream of the second Met and probably downstream of the H2 region (Fig. 2).

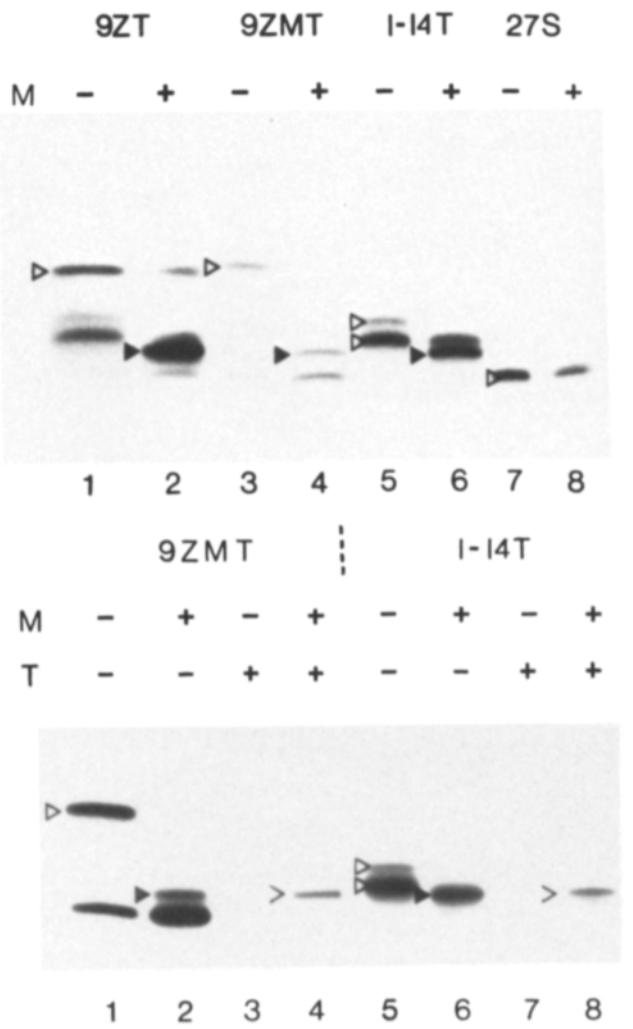

Figure 3. Identification of proteins translated from each initiation codon and their processing by microsomes. Genes 9ZT, 9ZMT, 1-14T, and 27S were inserted into the Bluescribe M13+ vector and transcribed using T7 RNA polymerase. RNAs were translated in rabbit reticulocyte lysates using [^{35}S]methionine (*top*) in the presence or absence of canine pancreatic microsomes (*M*) and (*bottom*) treated with trypsin (*T*) as indicated. Proteins analyzed on 10% polyacrylamide gels were detected by autoradiography. The unprocessed (▷) and processed (▶) product specific to each initiation codon is indicated in each case. The trypsin-resistant proteins are indicated (>).

Identification of the Major Cleavage Site in VP7

Empirical rules for predicting signal peptide cleavage sites in membrane proteins have been formulated from data accumulated for known cleavage events (24, 25) and these were used to examine VP7. Although residue 50 was predicted as the most likely cleavage site (Fig. 1), a previous study (4) suggested that cleavage occurred at a site between the two hydrophobic domains H1 and H2; two low-probability cleavage sites after residues 25 and 27 are predicted in this region (Fig. 1). Because Pro at certain positions in cleavage sites is particularly unfavorable (24), site-directed mutagenesis was used to introduce Pro at residues 23, 25, or 27. The respective genes were inserted into the SV40 expression vector pJC119 and transfected into COS cells in the presence of tunicamycin so that the size of the polypeptides produced could be assessed. In earlier work it was observed that failure to cleave the H1 hydrophobic domain resulted in an increase in the size of the protein (27). However, none of these mutations affected the size of VP7 suggesting that cleavage had still occurred in each case (data not shown). Similarly, protein 9ZMT carrying an additional Ser→Pro change at position 25 was also cleaved by microsomes in vitro (data not shown),

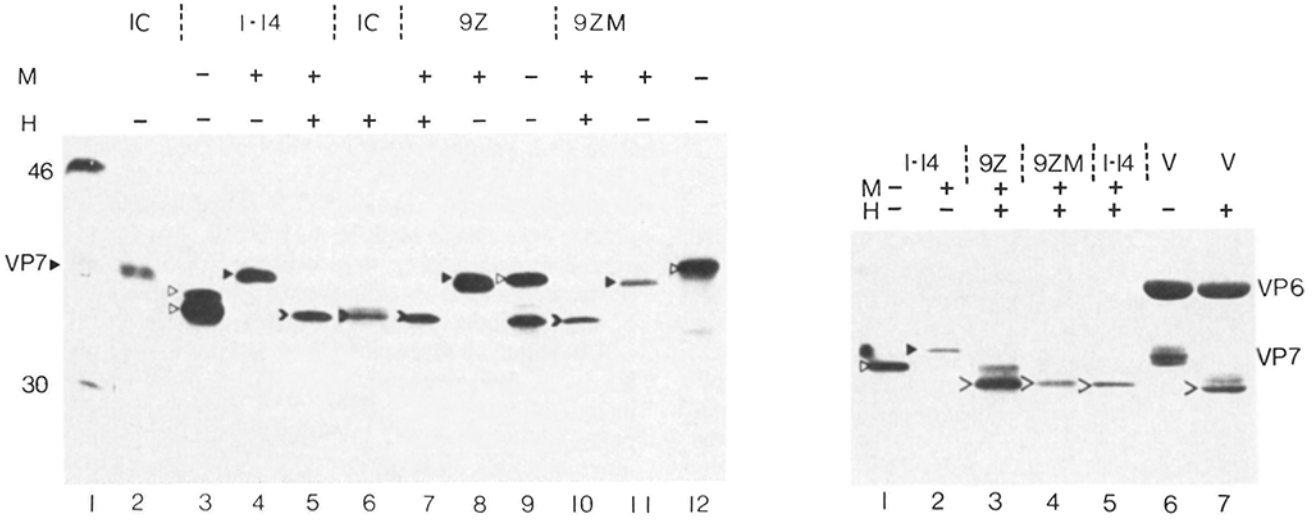

Figure 4. Comparison of VP7 proteins synthesized in vitro with VP7 present in (*left*) SA11 rotavirus-infected cells and (*right*) purified virus. Genes 9Z, 9ZM, and 1-14 in the Bluescribe vector were transcribed using T7 RNA polymerase and the RNAs translated in reticulocyte lysates in the presence and absence of microsomes (*M*). Samples were treated (+) with endo H (*H*) or not treated (−) as indicated. Proteins were immunoprecipitated from rotavirus-infected MA104 cells (*IC*) and similarly treated with endo H (lane *6*). (*V*) SA11 viral proteins. In vitro synthesized primary (▷), glycosylated (▶), and endo H–treated products (>) are indicated. In left panel, lane *1* contains SA11 VP7 (38 kD) and 46- and 30-kD proteins as markers.

despite the fact that this mutation was designed to prevent cleavage at both sites (24, 25).

The most favorable cleavage site Ala50↓Gln51 was therefore changed to Val50↓Gln51 (C3 mutation, Fig. 2, Table I), based on the observation that Val in the −1 position was also unfavorable for cleavage (24, 25) and had prevented cleavage of another mutant in vitro (27). RNAs transcribed in vitro from 9ZMTC3 and 1-14TC3 were translated in the presence and absence of microsomes and the products compared with those from the equivalent genes lacking the C3 mutation. Clearly this mutation (Ala→Val) inhibited microsome-dependent processing of the protein derived from the first AUG codon (Fig. 5, *left*, compare lanes *1* vs. *2* with *5* vs. *6*). The sensitivity of the proteins to trypsin digestion in the absence of microsomes (lanes *3* and *7*) and their resistance to trypsin when microsomes were added (lanes *4* and *8*) indicated that although the C3 mutation prevented processing of this polypeptide it did not prevent its entry into microsomes.

Similarly, the C3 mutation affected the processing of 1-14T translated from the second AUG (Fig. 5, *right*). In this case, the Ala→Val change did not completely prevent processing but resulted in the appearance of one, and perhaps two new products which were slightly smaller than processed 1-14T (compare lanes *2* and *6*) and resistant to trypsin digestion when microsomes were present (Fig. 5, *right*, lanes *4* and *8*). This suggested that when the primary cleavage site for signal peptidase was eliminated, a secondary, less favorable (unpredicted) site(s) was used. This alternative site(s) may also have been used in processing protein 9ZMTC3 in that similar trypsin-resistant bands were present in minor amounts in the presence of microsomes (Fig. 5, *left*, lane *8*).

Determination of the Partial Amino-terminal Sequence of VP7

We attempted to obtain amino-terminal sequence data for VP7 purified from virus but found the protein to be blocked, confirming the observation of other workers (1, 15). The putative cleavage site identified above predicts an amino-terminal glutamine residue for VP7 (Fig. 2), suggesting that the blocking group could be pyroglutamic acid. VP7 polypeptides translated and processed in vitro were therefore

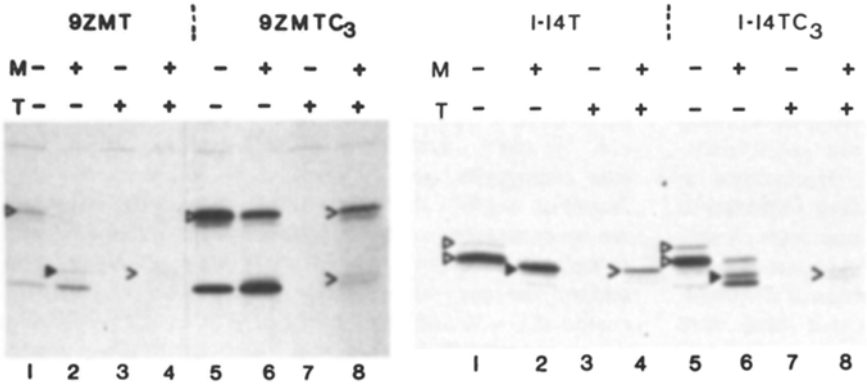

Figure 5. Effect of the C3 mutation on cleavage of nonglycosylated proteins translated from the first and second initiation codons. Genes 9ZMT and 9ZMTC3 (*left*) and 1-14T and 1-14TC3 (*right*) were transcribed in the Bluescribe system, translated in reticulocyte lysates in the presence and absence of microsomes (*M*), and treated with trypsin (*T*) as indicated. The unprocessed (▷), processed (▶), and trypsin resistant products (>) are indicated in each case.

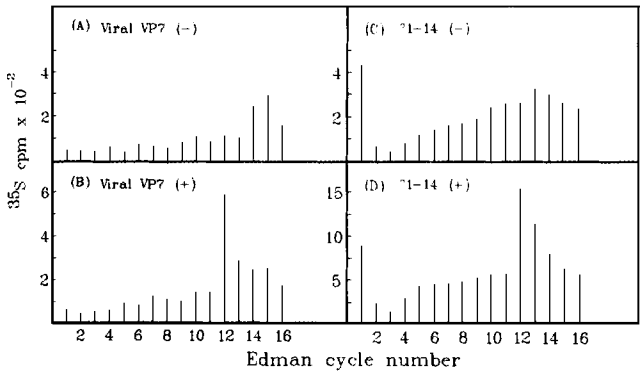

Figure 6. Partial amino-terminal sequence analysis of viral VP7 proteins and in vitro-translated 1-14 protein. [^{35}S]methionine-labeled viral VP7 proteins and 1-14 translated in vitro were separated by polyacrylamide gel electrophoresis and located by autoradiography. The proteins were excised, recovered by electroelution, and submitted to Edman degradation (*A*, *C*) before and (*B*, *D*) after digestion with pyroglutamate aminopeptidase. Radioactivity released at each cycle was monitored by liquid scintillation counting. In samples *A–D* there were 19,300, 35,000, 100,000, and 300,000 cpm of [^{35}S]methionine eluted from gels for analysis, respectively. There are 10 Met residues in VP7.

purified by gel electrophoresis and electroelution and subjected to Edman degradation before and after treatment with pyroglutamate aminopeptidase. Similar results were obtained for analysis of [^{35}S]methionine-labeled proteins 1-14 and 9ZM as well as nonglycosylated proteins 9ZMT and 1-14T (data not shown). In the absence of enzyme treatment no significant amount of radioactivity was released during 16 cycles of Edman degradation, except for cycle one (Fig. 6 *C*). This was not observed for 9ZMT (data not shown) and was probably due to contaminating, unprocessed 1-14 protein which had amino-terminal methionine (see above). After digestion with enzyme, which removes the blocked terminal residue, radioactivity was released at cycle 12 (Fig. 6 *D*), precisely the result expected if Gln51 were the amino-terminal residue in VP7. The result for 9ZMT was confirmed by analysis of [^{3}H]isoleucine-labeled protein (data not shown).

The [^{35}S]methionine-labeled VP7 proteins derived from purified virus (Fig. 4, *right*, lane *6*) were also analyzed. Without enzyme digestion, some radioactivity was reproducibly released at cycles 14 and 15 (Fig. 6 *A*). However, after enzyme treatment, most radioactivity was again released at cycle 12 (Fig. 6 *B*), indicating that Gln51 was also the amino-terminal residue in viral VP7.

To further confirm that cleavage occurred between Ala50 and Gln51, the amino-terminal glycopeptide of [^{3}H]mannose-labeled VP7 isolated from purified virus was released by digestion with *Staphylococcus aureus* V8 protease and isolated by HPLC. Such digestion of viral VP7 was expected to produce a peptide corresponding to amino acids Gln51-Glu74 (Fig. 2). However three peaks of [^{3}H]mannose-labeled material were recovered and these were subjected to amino acid analysis. The composition of one peak closely agreed with that expected for the amino-terminal peptide (data not shown). The elution pattern and amino acid compositions of the other two peaks suggested that these were shorter, related peptides, although their origin is unclear.

Discussion

Two in-frame initiation codons are conserved in all rotavirus VP7 genes so far sequenced and therefore the gene is potentially bicistronic (8, 9). We have identified proteins encoded from each possible initiation codon in the SA11 VP7 gene and examined their processing in vitro. Our results indicated that although the primary products translated from these codons were clearly distinct, the proteins cleaved and glycosylated by microsomes were indistinguishable in size. The processed proteins also migrated at a rate very similar to that of a polypeptide(s) present in rotavirus-infected cells.

A potential cleavage site downstream of the H2 region in VP7 (Fig. 2) was identified by scanning the amino acid sequence using rules compiled from cleavage of other signal peptides (24, 25). Through the introduction of a point mutation at this site, cleavage was largely eliminated and the use of the site was confirmed directly by partial amino-terminal sequence analysis. Radiolabeled VP7 proteins synthesized in vitro and the mixture of viral VP7 proteins yielded similar amino-terminal sequence profiles (Fig. 6). The determination of each partial sequence largely depended upon digestion of the proteins with pyroglutamate aminopeptidase. The amino acid composition of the amino-terminal glycopeptide purified from viral VP7 also supported the site directed mutagenesis data. Collectively the results indicated that processing in vivo and in vitro occurred in the same way, the major cleavage site lying between Ala50 and Gln51 (Fig. 2). This location contrasts with that proposed in another report in which it was suggested that processing occurred between the H1 and H2 domains (4). Our results demonstrate that neither H1 nor H2 are present in the mature form(s) of the protein.

Although Gln51 is the major amino terminus of VP7, there were also indications that the protein may have a "ragged" end, consistent with the presence of multiple protein species in the virus (Fig. 4, *right*, lanes *6* and *7*). In that only one methionine residue occurs near the proposed amino terminus (Met63, Fig. 2), the release of small but reproducible amounts of radioactivity at cycles 14 and 15 during Edman degradation (Fig. 6 *A*) suggested that the minor species may have unblocked amino termini beginning with Arg49 and Ala50. Similar results were also obtained for the analysis of processed proteins 9ZMT and 1-14T (data not shown) which were also heterogeneous (e.g. Fig. 5, *right*, lanes *2* and *4*). We interpret these data cautiously, however, because the composition of two glycopeptides purified from virus was more consistent with the loss of residues Gln51-Gly54 from the amino terminus. However, for the latter result we cannot rule out the possibility that partial, inappropriate cleavage by V8 protease produced these peptides, although we consider this to be unlikely.

The major cleavage site producing amino-terminal Gln appears to be a typical substrate for signal peptidase (Fig. 1) (24, 25) and is conserved in all VP7 serotypes (8, 9). However, cleavage by signal peptidase to produce amino-terminal Arg49 or Ala50 is much less likely (25) and these residues are not conserved in other VP7 proteins (8, 9). In the absence of the primary cleavage site in 1-14TC3, processing at a secondary site was observed. Recent data show that this occurred at Gly54 and to a lesser extent at Asn52, generating amino-terminal Ile and Tyr (Fig. 2) (Stirzaker, S. C., unpub-

lished results). These cleavage events were also not predicted by the computer program (Fig. 1) (25). Cleavage at an alternative site in yeast invertase was also observed after the introduction of an Ala→Val mutation in the −1 position (21). Inasmuch as we do not fully understand the origin of the minor proteins which may have amino-termini other than Gln, we do not exclude the possibility that VP7 processing may lack fidelity, or involve an enzyme other than signal peptidase.

Because the processed products translated from each AUG are the same, apparently the first initiation codon is not needed to produce VP7. This raises the question of why the codon is conserved in all the VP7 genes so far sequenced. There are two obvious alternative explanations. First, the codon is conserved because of some requirement for sequence conservation at the level of the RNA. The six VP7 gene sequences available show very high homology for the first ~80 nucleotides and bases 37–72 (which include the first AUG) are identical except for a single base change in one gene (8). Secondly, the codon may be used in vivo to produce low levels of a protein with a subtly different property, e.g., in its ability to interact with a membrane. Although the relative usage of each initiation codon in the infected cell is unknown, the simplest interpretation of the data is that VP7 is produced principally via the second initiation codon and second signal peptide region, and then cleaved between Ala50 and Gln51.

The principal phenomenon that must be explained in the light of the proposed cleavage site is the location of VP7 in the cell. Immunofluorescence studies using transfected cells (20) and analysis of the carbohydrate attached to VP7 (3) both indicate that the protein is not transported along the normal secretory pathway of the cell. Rather, VP7 inherently locates in the ER for subsequent assembly onto viral cores during infection. Further, after a 10-min pulse-labeling of proteins in the infected cell, very little uncleaved VP7 is detectable (5) suggesting that slow cleavage, akin to that seen for a mutant of yeast invertase (21), does not provide transient anchoring. However, VP7 does remain anchored after processing (10) and our data unequivocally show that cleavage occurs after Ala50, removing both of the hydrophobic domains. Clearly, therefore, only residues distal to residue 50 can be involved in retaining VP7 in the ER. Recently the tetrapeptide Lys-Asp-Glu-Leu has been proposed as a targeting signal for luminal ER proteins (16). However, this sequence is not present in VP7. The mechanism by which VP7 is retained in the ER is under further investigation.

We wish to thank L. Siegman for technical assistance, J. Back of the CSIRO Division of Food Research for her unstinting cooperation and technical assistance in performing the amino-terminal sequence analyses, C. Bucholtz for writing software for analysis of cleavage sites in proteins, W. F. Chadderton for competent technical assistance in preparing viral proteins for sequence analysis, and R. Hill and P. Jennings for comments on the manuscript.

This work was supported by grants from the World Health Organization, and the Medical Research Council of New Zealand. S. C. Stirzaker was supported by the CSIRO Division of Molecular Biology.

Received for publication 6 July 1987, and in revised form 22 September 1987.

References

1. Arias, C. F., S. Lopez, J. R. Bell, and J. H. Strauss. 1984. Primary structure of the neutralization gene of simian rotavirus SA11 as deduced from cDNA sequence. *J. Virol.* 50:657–661.
2. Both, G. W., J. S. Mattick, and A. R. Bellamy. 1983. Serotype-specific glycoprotein of simian 11 rotavirus: coding assignment and gene sequence. *Proc. Natl. Acad. Sci. USA.* 80:3091–3095.
3. Both, G. W., J. S. Mattick, L. J. Siegman, P. H. Atkinson, S. Weiss, A. R. Bellamy, J. E. Street, and P. Metcalf. 1983. Cloning of SA11 rotavirus genes: gene structure and polypeptide assignment for the type-specific glycoprotein. *In* The Double-stranded RNA Viruses. D. H. L. Bishop and R. W. Compans, editors. Elsevier Biomedical, New York. 73–82.
4. Chan, W.-K., M. E. Penaranda, S. E. Crawford, and M. K. Estes. 1986. Two glycoproteins are produced from the rotavirus neutralization gene. *Virology.* 151:243–252.
5. Ericson, B. L., D. Y. Graham, B. B. Mason, and M. K. Estes. 1982. Identification, synthesis, and modifications of simian rotavirus SA11 polypeptides in infected cells. *J. Virol.* 42:825–839.
6. Ericson, B. L., D. Y. Graham, B. B. Mason, H. H. Hansenn, and M. K. Estes. 1983. Two types of glycoprotein precursors are produced by the simian rotavirus SA11. *Virology.* 127:320–332.
7. Estes, M. K., E. L. Palmer, and J. F. Obijeski. 1983. Rotaviruses: a review. *Curr. Top. Microbiol. Immunol.* 105:123–184.
8. Gorziglia, M., Y. Aguirre, Y. Hoshino, J. Esparza, I. Blumentals, J. Askaa, M. Thompson, R. Glass, A. Z. Kapikian, and R. M. Chanock. 1986. VP7 serotype-specific glycoprotein of OSU porcine rotavirus: coding assignment and gene sequence. *J. Gen. Virol.* 67:2445–2454.
9. Gunn, P. G., F. Sato, K. F. H. Powell, A. R. Bellamy, J. R. Napier, D. R. K. Harding, W. S. Hancock, L. J. Siegman, and G. W. Both. 1985. Rotavirus neutralizing protein VP7: antigenic determinants investigated by sequence analysis and peptide synthesis. *J. Virol.* 54:791–797.
10. Kabcenell, A. K., and P. A. Atkinson. 1985. Processing of the rough endoplasmic reticulum membrane glycoproteins of rotavirus SA11. *J. Cell Biol.* 101:1270–1280.
11. Kapikian, A. Z., and R. M. Chanock. 1985. Rotaviruses. *In* Virology. B. N. Fields, editor. Raven Press, New York. p. 863–906.
12. Kozak, M. 1986. Bifunctional messenger RNAs in eukaryotes. *Cell.* 47: 481–483.
13. Kozak, M. 1986. Point mutations that define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell.* 44:283–292.
14. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (Lond.).* 227:680–685.
15. Lopez, S., C. F. Arias, J. R. Bell, J. H. Strauss, and R. T. Espejo. 1985. Primary structure of the cleavage site associated with trypsin enhancement of rotavirus SA11 infectivity. *Virology.* 144:11–19.
16. Munro, S., and H. R. B. Pelham. 1987. A C-terminal signal prevents secretion of luminal ER proteins. *Cell.* 48:899–907.
17. Nisbet, I. T., and M. W. Beilharz. 1985. Simplified DNA manipulations based on in vitro mutagenesis. *Gene Anal. Tech.* 2:23–29.
18. Petrie, B. L., H. B. Greenberg, D. Y. Graham, and M. K. Estes. 1984. Ultrastructural localization of rotavirus antigens using colloidal gold. *Virus Res.* 1:133–152.
19. Podell, D. N., and G. N. Abraham. 1978. A technique for the removal of pyroglutamic acid from the amino terminus of proteins using calf liver pyroglutamate amino peptidase. *Biochem. Biophys. Res. Commun.* 81: 176–185.
20. Poruchynsky, M. S., C. Tyndall, G. W. Both, F. Sato, A. R. Bellamy, and P. A. Atkinson. 1985. Deletions into an NH_2-terminal hydrophobic domain result in secretion of rotavirus VP7, a resident endoplasmic reticulum membrane glycoprotein. *J. Cell Biol.* 101:2199–2209.
21. Schauer, I., S. Emr, C. Gross, and R. Schekman. 1985. Invertase signal and mature sequence substitutions that delay intercompartmental transport of active enzyme. *J. Cell Biol.* 100:1664–1675.
22. Stearne, P. A., I. R. van Driel, B. Grego, R. J. Simpson, and J. W. Goding. 1985. The murine plasma cell antigen PC-1: purification and partial amino acid sequence. *J. Immunol.* 134:443–448.
23. Street, J. E., M. C. Croxson, W. F. Chadderton, and A. R. Bellamy. 1982. Sequence diversity of human rotavirus strains investigated by northern blot hybridization analysis. *J. Virol.* 43:369–378.
24. von Heijne, G. 1983. Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* 133:17–21.
25. von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 14:4683–4690.
26. Walter, P., R. Gilmore, and G. Blobel. 1984. Protein translocation across the endoplasmic reticulum. *Cell.* 38:5–8.
27. Whitfeld, P. L., C. Tyndall, S. C. Stirzaker, A. R. Bellamy, and G. W. Both. 1987. Location of signal sequences within the rotavirus SA11 glycoprotein VP7 which direct it to the endoplasmic reticulum. *Mol. Cell. Biol.* 7:2491–2497.
28. Wickner, W., and H. F. Lodish. 1985. Multiple mechanisms of protein insertion into and across membranes. *Science (Wash. DC).* 230:400–407.