The Genome-linked Protein of Picornaviruses. VIII. Complete Amino Acid Sequence of Poliovirus VPg and Carboxy-terminal Analysis of its Precursor, P3-9

By CHERYL J. ADLER,1 MARSHALL ELZINGA2 AND ECKARD WIMMER1.*

1 Department of Microbiology, School of Medicine, State University of New York at Stony Brook, Stony Brook, New York 11794 and 2 Department of Biology, Brookhaven National Laboratory, Upton, New York 11973, U.S.A.

(Accepted 17 August 1982)

SUMMARY

VPg, the genome-linked protein of poliovirus, and its putative precursor P3-9, were radiolabelled and subjected to carboxypeptidase-A digestion. The release of amino acids was followed by identification and quantification on an amino acid analyser. Both proteins were found to be co-terminal with a sequence of -valyl-glutamine-COOH, an observation that provides further evidence that host cell trimming of virus-specific peptides does not play a role in poliovirus protein processing. Radiolabelled VPg was subjected to automated Edman degradation. The combined results complete the structural analysis of VPg, a polypeptide 22 amino acids in length with a molecular weight of 2354. Only one form of VPg has been found linked to virion RNA and it originates by a cleavage at glutaminyl-glycine pairs at both termini. The observation is consistent with other cleavages found in the virus processing scheme.

INTRODUCTION

Poliovirus, a member of the Picornaviridae, is a small non-enveloped virus containing a single-strand RNA genome of plus polarity. The virus RNA genome (2.4 x 10^6 mol. wt.) is covalently linked to a small, basic protein, VPg (Lee et al., 1976, 1977) via an O4-(5'-uridylyl)-tyrosine (Ambros & Baltimore, 1978; Rothberg et al., 1978). This same protein has been shown to be linked to all newly formed poliovirus plus and minus strand RNAs as well as to nascent strands of the replicative intermediate (Nomoto et al., 1977; Flanegan et al., 1977; Pettersson et al., 1978; Wimmer, 1979).

We have previously reported a partial amino acid sequence for VPg based on radiochemical microsequence analysis (Kitamura et al., 1980). Here we present the complete amino acid sequence of VPg, including direct identification of its carboxy-terminus by carboxypeptidase-A analysis. Various laboratories had determined the molecular weight of VPg by gel filtration (Ambros & Baltimore, 1978) or SDS–polyacrylamide gel electrophoresis (SDS–PAGE) (Lee et al., 1977; Golini et al., 1978) to be 12000 or 5000 respectively; it is shown here to be 2354, a value corresponding to 22 amino acids. The aberrant migration cannot readily be explained by the primary sequence of VPg.

It has been reported for aphthovirus that two or more different forms of VPg are present on RNA (King et al., 1980). These two VPgs differ in their isoelectric points and in their amino acid compositions. Poliovirus RNA was also reported to have two forms of VPg with different charges (Richards et al., 1981) although the molecular basis of this difference was not determined. Our work, however, has failed to yield evidence for multiple forms of VPg.

The role of the genome-linked protein has not been established but it has been hypothesized to serve a role in the initiation of RNA replication (Nomoto et al., 1977; Pettersson et al., 1978; Wimmer, 1979), probably as a primer. Since VPg cannot be found free in the infected cell, Semler et al. (1982) have proposed that P3-9, a protein containing the VPg sequence, may be the
functional precursor protein, being cleaved at the moment of initiation to VPg. P3-9 was also analysed with carboxypeptidase-A and found to be co-terminal with VPg at -valyl-glutamine-COOH. Our data support the model that P3-9 may be the donor of VPg to the RNA chain. 

METHODS

**Virus and VPg RNA.** $^{32}$P-labelled poliovirus type 1 (Mahoney) was isolated from infected HeLa S3 suspension cultures as described previously (Lee et al., 1979). $^3$H and $^{14}$C amino acid labelling of virus and subsequent isolation were essentially the same as in Kitamura et al. (1980). Poliovirus VPg RNA was extracted from purified virus with phenol/chloroform/isoamyl alcohol (50:48:2, by vol.) and further purified through sucrose gradients. These RNA preparations do not contain detectable amounts of capsid proteins (Rothberg et al., 1978).

**P3-9 purification.** The membrane-associated non-structural protein P3-9 ([3H]valine- and [3H]glutamine-labelled) was obtained from the same infected cells from which the labelled virus was purified. Therefore, the specific activities of the labelled P3-9 and VPg were the same. At 7 h post-infection, cells were pelleted and the supernatant discarded. The pellets were suspended in 0-01 M-NaCl, 0-01 M-Tris-HCl pH 7-35, 0-0015 M-MgCl$_2$ (1/10 culture vol.), Dounced ten times and then centrifuged at 2000 g for 8 min. The supernatant was used to purify virus; the pellet was the source of P3-9. The pellet was dissolved in Laemmli sample buffer, boiled for 5 min, and electrophoresed on a 12-5% SDS-polyacrylamide gel (Laemmli, 1970). Using $^{35}$S]methionine-labelled intracellular virus proteins as markers, the protein bands were excised from the dried gel as described by Semler et al. (1981).

**Enzyme digestions.** VPg-pUp was released from virion RNA by RNase T2 (Calbiochem) (Lee et al., 1977; Nomoto et al., 1977), and then immediately subjected to automated Edman degradation. Digestion was at 37 °C for 30 min at an enzyme/RNA ratio of 2 units/40 μg RNA. VPg RNA and P3-9 were digested with carboxypeptidase-A (Worthington Biochemicals) following the procedure of Bhown et al. (1980) as modified by Emini et al. (1982). Briefly, the protein (still linked to RNA in the case of VPg) and carrier myoglobin were incubated for 10 min at 80 °C, cooled to room temperature and then digested at an enzyme/protein ratio of 1:4. The reaction took place at room temperature for the desired time and was terminated with two drops of glacial acetic acid, frozen immediately, and then lyophilized to dryness.

**Radiochemical microsequence analysis.** Sequence analysis was described earlier by Kitamura et al. (1980, and references therein) utilizing the Beckman Sequenator 890C and a Beckman Peptide Programme no. 102974. Degradation of all samples except $^{32}$P-labelled VPg-pUp was performed in the presence of polybrene. Scintillation counting was performed on one half of each butyl chloride extract except in the case of $^{32}$P-labelled VPg-pUp, where the entire extract was counted. High performance liquid chromatography (HPLC) analysis was performed on the remaining phenylthiohydantoin (PTH)-derivatized amino acids.

**Amino acid analysis.** Carboxypeptidase-A treated proteins were subjected to amino acid analysis and the liberated amino acids were identified and quantified. The lyophilized samples were dissolved in 0-5 ml buffer (0-2 M-sodium citrate pH 2-1 containing 15% polyethylene glycol) and run on a three buffer single column amino acid analyser (for details see Emini et al., 1982). Fractions (1.5 ml) were collected and assayed for radioactive content. The total counts in the peak fractions characteristic for a given amino acid (defined by comparing elution times for known standards) were used for quantifying that amino acid.

RESULTS

**Amino acid sequence analysis of VPg**

Earlier partial amino acid sequencing of radiolabelled VPg-pUp (Kitamura et al., 1980) mapped the protein into the poliovirus genome (see Fig. 1). A predicted (from the RNA) and confirming experimental amino acid sequence was generated. This sequence left the C-terminus undetermined, although labelling studies with $^{35}$S]methionine led us to conclude that the protein was at most 27 residues long. The tyrosine that links the protein to the RNA (Rothberg et al., 1978; Ambros & Baltimore, 1978) was predicted to be in position +3 but this could not be confirmed due to the use of polybrene in the sample cup of the sequenator. Positions +8 and +12 were also unconfirmed.

Fig. 1 shows the completed amino acid sequence of VPg. Using $^{32}$P-labelled VPg-pUp, without polybrene coating the cup, a distinct peak of radioactivity was found at position +3, an observation indicating the presence of tyrosine-pUp (Fig. 2a). Since one of the properties of VPg is that it adsorbs to glass surfaces, the polybrene proved unnecessary for sequencing. There
Fig. 1. Poliovirus protein processing scheme. The polyprotein NCVPOO is divided into three regions P1, P2 and P3 which are the result of nascent cleavages. The P1 region contains the capsid proteins, P2 has non-structural proteins of unknown function and the P3 region contains replication proteins as well as a proteinase. The amino acid sequences of P3-9 and VPg as well as their surrounding sequences are shown below. The sequence of P3-9 has been in part deduced from the RNA and in part by amino acid sequence analysis with 16 positions confirmed (Semler et al., 1982; B. L. Semler & C. W. Anderson, unpublished results). The arrows represent cleavage sites for the virus-encoded proteinase.
was no washout of the peptide prior to step 3, and tyrosine-pUp did not stick to the spinning cup as it had in the presence of polybrene.

Edman degradation of [3H]asparagine-labelled VPg-pUp gave the results seen in Fig. 2(b). Two peaks of radioactivity appeared at steps 8 and 12. These were confirmed as being asparagine by HPLC analysis. This completed the amino acid sequence from position +1 to +21. The RNA sequence predicted glycine residues at positions +1, +5 and +23 but Edman degradation of [3H]glycine-labelled VPg-pUp gave peaks in positions +1 and +5, but not for position +23 (data not shown). Our failure to find glycine at +23 could have been explained by low yields (washout) of the remaining small polypeptide in the spinning cup; alternatively, the peptide was smaller than 23 amino acids. Labelling with [14C]glutamine and subsequent Edman degradation gave the result seen in Fig. 2(c). It appeared that glutamine was in position +22, but due to low yield the result seemed tenuous.

**Carboxy-terminal analysis of VPg**

Due to the inability of radiolabelled automated Edman degradation to confirm unequivocally the presence of glutamine at position +22, and given our suspicions that this was the carboxy-terminus of VPg, we undertook carboxypeptidase-A analysis of VPg. The analysis was performed with radiolabelled proteins rather than unlabelled, because sufficient amounts of the protein cannot be prepared from virions and unbound VPg does not exist in the infected cell (Semler et al., 1982).

As outlined in Methods, released amino acids were identified by an amino acid analyser. The percentage of a released amino acid was quantified by comparing the total counts in the identified peak with the calculated counts expected from the complete release of a single residue. The latter value was calculated from the total counts incorporated into the protein and the number of residues of the particular amino acid in the protein. This procedure assumes equivalent incorporation of an amino acid within the entire protein.

Fig. 3(a) shows the kinetics of release of glutamine and valine from the C-terminus of VPg.
**VPg sequence and C-terminus of its precursor**

![Graph](image)

**Fig. 3.** Release of [3H]glutamine (■) and [3H]valine (○) from carboxypeptidase-A-treated proteins. (a) VPg; (b) P3-9. Each set of reactions was carried out on identical protein samples. Reactions were terminated at 5, 10, 30 and 60 min after the addition of carboxypeptidase-A to the protein samples (100% expected release for each residue represents 2000 cts/min).

**Table 1.** Release of [3H]-labelled amino acids by carboxypeptidase-A*

<table>
<thead>
<tr>
<th>Protein</th>
<th>% Release of cts/min expected for one residue</th>
<th>Reaction time (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[3H]GLy VPg</td>
<td>&lt;0.1</td>
<td>2</td>
</tr>
<tr>
<td>[3H]Gln P3-9</td>
<td>77.8</td>
<td>4</td>
</tr>
</tbody>
</table>

* 100% expected release for each residue represents 2000 cts/min.

The fast exponential release of glutamine is highly characteristic of a carboxy-terminal amino acid, while the slower release of valine is characteristic of a penultimate amino acid (Ambler, 1967). These results confirmed –valyl-glutamine-COOH as the carboxy-terminus of VPg. Digestion of [3H]glycine VPg (Table 1) gave no glycine peak after 2 h. The counts were recovered in undigested protein. The 72% release of glutamine (Fig. 3a) at 60 min as opposed to complete release may be due to partial inhibition of the enzyme by SDS.

**Carboxy-terminal analysis of P3-9**

P3-9, a virus-specific polypeptide, was shown to contain the sequence of VPg (Semler et al., 1982) by immunoprecipitation with antibodies directed to the carboxy-terminus of VPg. It seemed likely from sequence, molecular weight and peptide fragmentation studies (Semler et al., 1982; Kitamura et al., 1981) that the proteins were co-terminal at –valyl-glutamine–COOH.

Therefore, [3H]glutamine- and [3H]valine-labelled P3-9 were subjected to carboxypeptidase-A analysis under the same conditions used for VPg. The curves generated in these experiments are shown in Fig. 3(b). The kinetics are similar to those seen in Fig. 3(a), indicating glutamine as the terminal amino acid and valine as the penultimate. The greater overall release of valine from P3-9 as compared to VPg may be due to it having been treated in a harsher (more denaturing) manner prior to digestion (see Methods).

Examination of the amino acid sequence of P3-9 (Fig. 1) reveals another valyl–glutamine pair at position + 50/+ 51. These amino acids are preceded by another glutamine at position + 49. In order to preclude this from being the carboxy-terminus of P3-9, a longer digestion of [3H]-glutamine P3-9 was performed. The results are seen in Table 1. We therefore conclude that this valyl–glutamine pair at + 50/+ 51 cannot be a termination site for P3-9, an observation in agree-
ment with sequencing studies. The 77.8% release of cts/min expected for one residue is just slightly higher than the value obtained for a 1 h digestion, and is consistent with the presence of only one glutamine.

**DISCUSSION**

In the results presented here we have conclusively shown the complete primary structure of poliovirus VPg. It is 22 amino acids in length, 2354 Mr, and terminates in –valyl-glutamine–COOH. The nature of the carboxy-terminus did not surprise us since in the poliovirus polyprotein, from which VPg is ultimately derived, the terminal glutamine of VPg is followed by a glycine. Recently, glutaminyl-glycine pairs have been shown to be cleavage sites for a virus-specific proteinase (Hanecak et al., 1982). Moreover, the glycine residue following the carboxyterminal glutamine of VPg is the amino terminus of P3-2, the cleavage product adjacent to VPg (Semler et al., 1981). Nevertheless, with this indirect evidence alone, one could not just assume that the carboxy-terminus of VPg is glutamine, particularly since the biosynthesis of VPg is peculiar and appears to be regulated by mechanisms involved in RNA synthesis (Wimmer, 1982). Moreover, there are four glutaminyl-glycine bonds in the polyprotein that remain uncleaved for unknown reasons. Our results are consistent with other poliovirus proteins whose carboxy-termini have been examined (Emini et al., 1982), in that no carboxy-terminal ‘trimming’ has been observed, once the glutaminyl-glycine bond has been cleaved.

The tyrosine residue which forms the linkage to the genome RNA (Rothberg et al., 1978; Ambros & Baltimore, 1978) is at position +3. VPg is basic, does not precipitate in 5 to 20% trichloroacetic acid or 10% acetic acid, adsorbs rapidly to glass or plastic surfaces and is insoluble in water unless detergent is present (these properties and others are reviewed by Adler et al., 1981), although Baron & Baltimore (1982) have recently reported that chemically synthesized VPg is water-soluble. The primary sequence cannot as yet be made to explain these properties, or the behaviour of VPg on SDS–PAGE or gel filtration, where the protein gives a false molecular weight estimate.

Only one form of VPg was found in this and past (Kitamura et al., 1980) sequencing studies. Isoelectric focusing of VPg-pUp, that was removed from virion RNA with RNase T2, also failed to separate VPg into more than one form (C. J. Adler & E. Wimmer, unpublished results). There was at no time two labels in any fraction from the sequenator, as all amino acids except tyrosine at position +3 were confirmed by HPLC analysis. It is possible that the two forms seen by Richards et al. (1981) may have arisen from post-translational modification of VPg. It is noteworthy that poliovirus has only one coding region for VPg, whereas aphthovirus RNA appears to code for more than one species of VPg that map in tandem (Forss & Schaller, personal communications).

VPg has never been found free in infected cells and it has been speculated that this is due to the cleavage of VPg from a precursor, P3-9, at the moment of initiation of RNA synthesis (Semler et al., 1982 and references therein). The observation reported here, that VPg and P3-9 are co-terminal at their carboxy ends, supports this hypothesis. It is interesting that P3-9 is a membrane-bound protein (Semler et al., 1982) and we believe that a segment of hydrophobic amino acids (residues +59 to +80, see Fig. 1) preceding the basic carboxy-terminal VPg tract (residues +88 to +109) is responsible for this property. If P3-9 is the membrane-bound donor for VPg, then a single cut at a glutaminyl–glycine bond (residues +87/+88) would release VPg. Since intracellular P3-9 is a stable protein, but intracellular unbound VPg is not (Dorner et al., 1981; Sangar et al., 1981), one could speculate that the linking of the first nucleotide of the RNA chain may be due to the O4-hydroxy group of tyrosine at residue +90 of P3-9. This linking could then make P3-9 a substrate for proteolytic cleavage, possibly by a conformational change, liberating VPg-pU...
REFERENCES


GOLINI, F., NOMOTO, A. & WIMMER, E. (1978). The genome-linked protein of picornaviruses. IV. Difference in the VPg's of encephalomyocarditis virus and poliovirus as evidence that the genome-linked proteins are virus-coded. Virology 89, 112–118.


LEE, Y. F., NOMOTO, A. & WIMMER, E. (1976). The genome of poliovirus is an exceptional eukaryotic mRNA. Progress in Nucleic Acid Research and Molecular Biology 19, 89–95.


(Received 28 March 1982)