Carboxy-Terminal Analysis of Poliovirus Proteins: Termination of Poliovirus RNA Translation and Location of Unique Poliovirus Polyprotein Cleavage Sites

EMILIO A. EMINI,^{1*} MARSHALL ELZINGA,² AND ECKARD WIMMER¹

Department of Microbiology, School of Medicine, State University of New York at Stony Brook, Stony Brook, New York 11794,¹ and Department of Biology, Brookhaven National Laboratory, Upton, New York 11973²

Received 3 November 1981/Accepted 3 December 1981

The carboxy-terminal amino acids of a number of poliovirus proteins were determined by carboxypeptidase A analysis. The nonstructural proteins P3-2, P3-4b and their precursor, P3-1b, were found to be coterminal with a sequence of -Ser-Phe-COOH. As these proteins are coded for at the extreme 3' end of the viral RNA, it is possible to establish the termination site of translation at nucleotide 7,361, 73 nucleotides before the start of the polyadenylic acid tract of the RNA. Two additional nonstructural proteins, P2-X and its precursor, P2-3b, were also found to be coterminal with a sequence of -Phe-Gln-COOH. This result confirms the existence of at least one Gln-Gly proteolytic cleavage site. These Gln-Gly cleavage sites are predicted from the nucleotide sequence to be ubiquitous throughout the poliovirus genome. The only exceptions are the cleavage sites at the carboxy termini of the structural proteins VP4 and VP1. Carboxypeptidase A analysis of VP1 establishes a terminal sequence of -Thr-Tyr-COOH, and similar analysis of VP4 shows Asn to be the terminal amino acid residue, observations that prove the existence of the exceptional C-terminal amino acids. In none of the analyzed cases has C-terminal trimming after cleavage been observed.

Poliovirus, a member of the Picornaviridae family of animal viruses, contains as its genome a single-stranded RNA (2.4×10^6 daltons) which acts as mRNA upon host cell infection (10). The RNA is covalently attached at its 5' end to a small protein (17-19), and its 3' end is polyadenylated (3, 26).

Recently, the complete nucleotide sequence of the genomic RNA of a poliovirus type 1 (Mahoney) strain was determined (11, 12, 20). The RNA contains a single open reading frame which spans 89% of its length. Upon translation in the infected host cell, the viral mRNA yields a large polyprotein (NCVPOO) which represents the RNA's total coding capacity. This polyprotein is post-translationally cleaved to produce, via a number of intermediate precursor polypeptides, the virus' structural and nonstructural proteins (Fig. 1) (8, 9, 24). The exact coding location of each of the proteins on the mRNA was established by comparing the amino-terminal amino acid sequence of each protein with the nucleotide codon sequences in the open reading frame (11, 15, 22, 24). Analysis of the RNA's nucleotide sequence and of the proteins' predicted amino acid sequences yields the following tentative conclusions.

(i) The termination site of translation of the

viral mRNA occurs at nucleotide 7,361, 73 nucleotides before the start of the polyadenylic acid tract. Several nonstructural proteins are predicted to terminate at this site: P3-2, P3-4b, and their precursor, P3-1b.

(ii) Viral protein cleavages occur at specific sites characterized by the presence of Gln-Gly pairs. Available evidence suggests that a virus-specific protease(s) is responsible for breaking the peptide bond between these two amino acids, yielding proteins with Gly at the amino terminus and Gln at the carboxy terminus (for references, see reference 11). The only exceptions are the cleavages between the viral structural proteins, VP4 and VP2, and between VP1 and P2-3b. The former apparently occurs at an Asn-Ser site, and the latter occurs at a Tyr-Gly site. The cleavage between VP4 and VP2 is functionally different from the other viral protein cleavages. It occurs at viral RNA encapsidation and virion maturation (reference 21 and literature cited therein). However, the substitution of Tyr for Gln at the carboxy terminus of VP1 cannot be accounted for in functional terms.

The goal of this study was to determine the carboxy-terminal amino acid sequences of four polioviral proteins: VP4, VP1, P2-X, and P3-2. The carboxy terminus of P3-2 provided evidence

FIG. 1. Poliovirus protein processing pathways. The viral RNA is denoted by a heavy line, and the proteins are denoted by finer lines. The figure is not drawn to scale with respect to relative protein sizes. P1, P2, and P3 represent the three main cleavage regions of the viral polyprotein. The proteins denoted by the enlarged letters and numbers were used for carboxy-terminal amino acid analysis in the studies reported here. The numbers in parentheses are the molecular weights ($\times 1,000$) of each of the individual proteins as calculated from their predicted amino acid sequences (11). The numbers on the RNA genome represent the exact sites at which some proteolytic cleavages of the protein products occur as established from data presented in this paper and in reference 22. The exact site of the termination of translation is also indicated (TERM.)

for the exact termination site of translation on the viral mRNA. The terminus of P2-X confirmed the existence of Gln-Gly cleavage sites, whereas the termini of the two structural proteins confirmed the nucleotide sequences which predict unique cleavage sites between VP4 and VP2 and between VP1 and P2-3b. In addition, evidence was provided for the coterminalization of P3-2, P3-4b, and their precursor, P3-1b, as well as for coterminalization of P2-X and its precursor, P2-3b.

MATERIALS AND METHODS

Labeling and purification of viral proteins. (i) **Non-structural proteins (P3-2, P3-4b, P3-1b, P2-X, P2-3b).** Approximately 3.0×10^7 HeLa S3 cells were infected with poliovirus type 1 (Mahoney) at a multiplicity of infection of 50 PFU/cell. At 2.5 h postinfection, 1.0 to 2.0 mCi of the ^3H -labeled amino acid in question was added in a small amount of medium. At 5.0 h postinfection, the infected cells were harvested by pelleting, washing with phosphate-buffered saline, and pelleting again. The final pellet was resuspended in Laemmli sample buffer (62.5 mM Tris [pH 6.8], 10% glycerol, 2.0% sodium dodecyl sulfate [SDS], 5.0% 2-mercaptoethanol, 0.005% bromophenol blue).

A greater yield of the precursor proteins (P3-1b and P2-3b) was obtained by treating the infected cells at 2.75 h postinfection with 0.8 mM ZnCl_2 , a specific inhibitor of the viral proteolytic reactions (6). In these cases, labeling was carried out from 3.0 to 4.0 h postinfection, followed by harvesting of the cells as described above.

Protein purification was carried out by subjecting the preparation to SDS-polyacrylamide gel electrophoresis in a 12.5% Laemmli gel (14). With [^{35}S]methionine-labeled viral proteins as markers, the protein bands of interest were excised from the dried gel. The gel pieces were rehydrated, and the protein was elec-

troeluted from the gel by an ISCO 1750 sample concentrator with SDS-free buffers (0.05 M NH_4HCO_3 in the outer chamber and 0.01 M NH_4HCO_3 in the sample compartment) containing 100 μg of myoglobin in the sample compartment as carrier. The eluted protein was lyophilized, dissolved in 0.2 M *N*-ethylmorpholine acetate (Pierce Chemical Co.) (pH 8.5)–0.1% SDS and stored at -20°C until used.

(ii) **Structural proteins (VP4, VP1).** HeLa S3 cells were infected as above, except that labeling with the ^3H -labeled amino acid was from 2.5 to 7.0 h postinfection. The cells were harvested as above, suspended in lysis buffer (0.01 M NaCl, 0.01 M Tris [pH 7.35], 1.5 mM MgCl_2) and lysed by several cycles of freeze-thawing. The cell nuclei were pelleted. Virus was then pelleted from the supernatant by spinning at $80,000 \times g$ for 5 h in 1.0% SDS. The virus was suspended in 0.1 buffer (0.1 M NaCl, 0.01 M Tris [pH 7.5], 1.0 mM EDTA) and purified by velocity sedimentation through a gradient of 15 to 30% (wt/wt) sucrose–0.1 buffer and 0.5% SDS. The viral band was collected, and the virus was pelleted as described above and then suspended in Laemmli sample buffer. The virion components were dissociated by heating to 100°C for 2 min.

Protein purification was carried out exactly as outlined above.

Carboxypeptidase A analysis of carboxy-terminal amino acids. Carboxypeptidase A was obtained from Worthington Diagnostics. The procedure used was basically that of Bhowan et al. (4). Briefly, the protein in 0.2 M *N*-ethylmorpholine acetate (pH 8.5)–0.1% SDS was placed at 80°C for 10 min. After cooling to room temperature, carboxypeptidase A in 0.2 M *N*-ethylmorpholine acetate (pH 8.5) was added at an enzyme/protein ratio of 1:4. (The amount of protein involved was essentially that of the carrier myoglobin.) The reaction was allowed to proceed at room temperature for the desired time and then terminated by the addition of 2 drops of glacial acetic acid. The sample was immediately lyophilized.

The liberated amino acids were conclusively identi-

FIG. 2. (A) Release of [3 H]Phe (●) and [3 H]Ser (○) from carboxypeptidase A-treated P3-2. (B) Release of [3 H]Gln (●) and [3 H]Phe (○) from carboxypeptidase A-treated P2-X. (C) Release of [3 H]Asn (●) from carboxypeptidase A-treated VP4. (D) Release of [3 H]Tyr (●) and [3 H]Thr (○) from carboxypeptidase

fied and quantitated by amino acid analysis. The lyophilized samples were dissolved in 0.5 ml of buffer (0.2 M sodium citrate [pH 2.1] containing 15% polyethylene glycol) and run on a three-buffer, single-column amino acid analyzer with the ninhydrin pump turned off. The first two buffers were those described earlier (24), and the third was 1.5 M sodium–0.1 M citrate (pH 4.6). The column was 22 by 0.6 cm, the resin was Bio-Rad Aminex A-7, the buffer flow rate was 20 ml/h, and the temperature was 54°C. Fractions (0.5 ml) were collected every 1.5 min and assayed for radioactive counts. The total counts in the peak fractions characteristic for a given amino acid (determined by comparing to elution times for known standards) were used for quantitating that amino acid (see below).

RESULTS

The strategy employed in these experiments was dictated by the difficulty encountered in obtaining enough highly purified virus-specific proteins to detect, by UV light absorbance, the release of amino acids after treatment of the protein with carboxypeptidase. Hence, the protein to be analyzed was radiolabeled with the specific amino acids predicted, from the nucleotide sequence of the viral RNA, to be at the carboxy terminus. Since poliovirus effectively turns off host cell protein synthesis (reviewed in reference 7), radiolabeled amino acids are incorporated only in viral proteins. Individual preparations were made for each amino acid. The labeled protein was then purified from other labeled viral proteins and from the bulk of nonlabeled host cell proteins. The purified protein was subjected to carboxypeptidase A treatment for the specified times, and the released amino acids were identified by an amino acid analyzer. The percentage of released, free amino acid was quantitated by comparing the total radioactive counts in the amino acid peak from the analyzer with the calculated counts expected from complete release of a single residue. The latter value was calculated from the total counts incorporated in the purified protein and the total number of residues of the amino acid in question predicted to be in the protein by the nucleotide sequence. This procedure, of course, assumes equivalent incorporation of an amino acid within the entire protein.

Termination site of translation. The kinetics of release of Phe and Ser from the carboxy terminus of P3-2 are shown in Fig. 2A. The fast, exponential release of Phe is highly characteristic of a C-terminal amino acid, whereas the

A-treated VP1. Proteins were labeled with a single amino acid at a time. Each set of reactions was carried out on identical protein samples. Reactions were terminated at 5, 10, 15, and 60 min after the addition of the carboxypeptidase to the protein samples.

slower release of Ser is characteristic of a penultimate amino acid residue (1). As predicted by the nucleotide sequence, these results confirm a carboxy terminus for P3-2 of -Ser-Phe-COOH. It should be noted that the very slow release of Ser seen here is probably due to the somewhat refractory nature of this amino acid to carboxypeptidase A. Furthermore, the less than complete release of both amino acids by 60 min may be due to a minor inhibition of carboxypeptidase action by the presence of SDS in the reaction buffers. Nevertheless, the relative rates of release of the amino acids support the prediction of the nucleotide sequence.

If the translation is shifted from the predicted reading frame by one or two nucleotides, the new amino acid sequences of P3-2 predict the presence of an Ile residue at or near the carboxy terminus. Hence, [^3H]Ile-labeled P3-2 was digested with carboxypeptidase A, and the release of Ile was monitored. This was done to eliminate the possibility that the predicted termination site might be wrong due to a nucleotide sequencing error involving the insertion or deletion of one or two nucleotides near the 3' end of the RNA. No Ile was released after 90 min of carboxypeptidase A treatment (data not shown).

Release of more than one Phe or Ser from the carboxy terminus of P3-2, even at prolonged times of incubation, would not be expected because of two Arg residues at positions -6 and -7 of the polypeptide (Fig. 3). These arginines, similar to lysine residues, block further degradation of P3-2 by the exopeptidase (2); the next Phe in P3-2 occurs only at residue -30 (11); the next serine occurs at residue -11 (Fig. 3). Similarly, Ile residues would not be expected to be released as the first Ile occurs at position -20.

The cleavage scheme of the poliovirus polypeptide predicts that polypeptides P3-1b, P3-2, and P3-4b share identical amino acid sequences (Fig. 1) (11 and 22). We therefore labeled P3-1b and P3-4b with [^3H]Phe or [^3H]Ser. The polypeptides were carboxypeptidase A treated for 30 min each, and the release of the labeled amino acids was quantitated. These results (Table 1) show that approximately equivalent relative amounts of Phe and Ser are released by all three proteins. This indicates that the three proteins are coterminal at the carboxy ends and that carboxy-terminal trimming does not occur during processing from precursor to products.

Carboxy terminus of P2-X. The kinetics of release of Gln and Phe from the carboxy terminus of P2-X is shown in Fig. 2B. As predicted by the nucleotide sequence, only one Gln and one Phe are expected to be released because of two Arg residues at -13 and -14 blocking further digestion. The results confirm a terminus for

P2-X of -Phe-Gln-COOH. Since Gly is known to be present at the amino terminus of the following protein, P3-1b (22), the existence of at least one Gln-Gly proteolytic cleavage site has been proven.

In addition, [^3H]Phe- and [^3H]Gln-labeled P2-3b, the precursor protein to P2-X, was carboxypeptidase treated for 30 min; 49.7% of the [^3H]Gln counts per minute per residue and 37.3% of the [^3H]Phe counts per minute per residue were released. These results show that P2-X and P2-3b are coterminal at their carboxy ends and that P2-3b is not trimmed from this end while being processed to P2-X.

Unique cleavage sites of VP4 and VP1. Of particular interest to us are the cleavage sites between capsid proteins VP4 and VP1 and between capsid protein VP1 and the nonstructural polypeptide P2-3b. As predicted by the nucleotide sequence, these sites are Asn-Ser and Tyr-Gly, respectively, and differ from all other known Gln-Gly cleavage sites (11, 15, 22, 23).

The kinetics of release of Asn from the carboxy terminus of VP4 is presented in Fig. 2C. The result clearly confirms the prediction of Asn as the carboxy-terminal amino acid of VP4 and provides supporting evidence for the unique proteolytic cleavage site between VP4 and VP2.

Similarly, VP1 labeled with [^3H]Tyr or [^3H]Thr was subjected to carboxypeptidase A digestion. The kinetics of release of Tyr and Thr from VP1 is shown in Fig. 2D. The results conform with the amino acid sequence predicted from the nucleotide sequence and establish a carboxy terminus for VP1 of -Thr-Tyr-COOH.

Since the unique cleavage site between VP1 and P2-3b appears to have no functional significance, a further test for the presence of Tyr instead of Gln at this cleavage site was performed. [^3H]Gln-labeled VP1 was treated with carboxypeptidase A for 30 min. No release of Gln was noted (data not shown).

As can be seen from Fig. 3 basic amino acids near the C terminus prevent release of a second molecule of the respective labeled amino acids in VP4 as well as in VP1.

DISCUSSION

Polynucleotide sequence analysis is subject to errors no matter how carefully the work is carried out or what method is used. The amino acid sequences predicted from the nucleotide sequence must therefore be considered as tentative, and they should be verified, at least for areas of special interest, by amino acid analysis.

The C-terminal amino acids of P3-1b, P3-2, and P3-4b presented here prove that termination of viral translation occurs at . . . Ser-Phe, 73 nucleotides before the start of the poly(A) tail of the RNA. Thus, in poliovirus translation the

a common phenomenon. Cleavages of the hemagglutinin glycoprotein of influenza virus by host cellular proteinase(s) involves the removal of a pentapeptide in the case of fowl plaque virus and of a single amino acid in all other influenza virus strains analyzed (16). Similarly, the *gag* gene product of Rous sarcoma virus is processed by a viral proteinase to four polypeptides where-by nine amino acids are removed between products p27 and p12, and two amino acids are removed between p19 and p10 (E. Hunter, personal communication). In contrast, the cleavages reported here and the cleavage between VPg and P3-2 (22) do not result in any trimming of termini. Absence of trimming has also been observed in the proteolytic processing of the aphthovirus (foot-and-mouth disease virus) capsid polypeptides (5, 13).

ACKNOWLEDGMENTS

We thank Nicholas Alonzo for invaluable technical assistance.

This work was supported by Public Health Services grants AI-15122 and CA-28146 from the National Institutes of Health and by the U.S. Department of Energy. E.A.E. is a postdoctoral fellow of the American Cancer Society.

LITERATURE CITED

- Ambler, R. P. 1967. Enzymic hydrolysis with carboxypeptidases. *Methods Enzymol.* 11:155-166.
- Ambler, R. P. 1967. Carboxypeptidases A and B. *Methods Enzymol.* 11:436-445.
- Armstrong, J. A., M. Edmonds, H. Nakazato, B. A. Phillips, and M. H. Vaughan. 1972. Polyadenylic acid sequences in the virion RNA of poliovirus and eastern equine encephalitis virus. *Science* 176:526-528.
- Bhown, A. S., J. C. Bennett, and E. Hunter. 1980. Alignment of the peptides derived from acid catalyzed cleavage of an aspartylprolyl bond in the major internal structural polypeptide of avian retroviruses. *J. Biol. Chem.* 255:6962-6965.
- Boothroyd, J. C., P. E. Highfield, G. A. M. Cross, D. J. Rowlands, P. A. Lowe, F. Brown, and T. J. Harris. 1981. Molecular cloning of foot and mouth disease virus genome and nucleotide sequences in the structural protein genes. *Nature (London)* 290:800-802.
- Butterworth, B. E., and B. D. Korant. 1974. Characteristics of the large picornaviral polypeptides produced in the presence of zinc ion. *J. Virol.* 14:282-291.
- Ehrenfeld, E. 1979. *In vitro* translation of picornavirus RNA, p. 175-190. *In* R. Perez-Bercoff (ed.), *The molecular biology of picornaviruses*. Plenum Publishing Corp., New York.
- Holland, J. J., and E. D. Kiehn. 1968. Specific cleavage of viral proteins as steps in the synthesis and maturation of enteroviruses. *Proc. Natl. Acad. Sci. U.S.A.* 60:1015-1022.
- Jacobson, M. F., and D. Baltimore. 1968. Polypeptide cleavages in the formation of poliovirus proteins. *Proc. Natl. Acad. Sci. U.S.A.* 61:77-84.
- Kitamura, N., C. Adler, and E. Wimmer. 1980. Structure and expression of the picornavirus genome. *Ann. N.Y. Acad. Sci.* 345:183-201.
- Kitamura, N., B. L. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, J. J. Lee, S. van der Werf, C. W. Anderson, and E. Wimmer. 1981. Primary structure, gene organization, and polypeptide expression of poliovirus RNA. *Nature (London)* 291:547-553.
- Kitamura, N., and E. Wimmer. 1980. Sequence of 1,060 3'-terminal nucleotides of poliovirus RNA as determined by a modification of the dideoxynucleotide method. *Proc. Natl. Acad. Sci. U.S.A.* 77:3196-3200.
- Kupper, H., W. Keller, C. Kurz, S. Forss, H. Schaller, R. Franze, K. Strohmaier, O. Marquardt, V. G. Zaslavsky, and P. H. Hofschneider. 1981. Cloning of cDNA of major antigen of foot-and-mouth disease virus and expression in *E. coli*. *Nature (London)* 289:555-559.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* 227:681-685.
- Larsen, G. R., C. W. Anderson, A. J. Dorner, B. L. Semler, and E. Wimmer. 1982. Cleavage site within the poliovirus capsid protein precursors. *J. Virol.* 41:340-344.
- Laver, G., and G. Ails. 1980. *Structure and variation in influenza virus*. Elsevier North-Holland, New York.
- Lee, Y. F., A. Nomoto, B. M. Detjen, and E. Wimmer. 1977. The genome-linked protein of picornaviruses. I. A protein covalently linked to poliovirus genome RNA. *Proc. Natl. Acad. Sci. U.S.A.* 74:59-63.
- Lee, Y. F., A. Nomoto, and E. Wimmer. 1976. The genome of poliovirus is an exceptional eukaryotic mRNA. *Prog. Nucleic Acid Res. Mol. Biol.* 19:89-95.
- Nomoto, A., N. Kitamura, F. Golini, and E. Wimmer. 1977. The 5'-terminal structures of poliovirion RNA and poliovirus mRNA differ only in the genome-linked protein VPg. *Proc. Natl. Acad. Sci. U.S.A.* 74:5345-5349.
- Racaniello, V. R., and D. Baltimore. 1981. Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. *Proc. Natl. Acad. Sci. U.S.A.* 78:4887-4891.
- Rueckert, R. R. 1976. On the structure and morphogenesis of picornaviruses, p. 131-213. *In* H. Fraenkel-Conrat and R. R. Wagner (ed.), *Comprehensive virology*, vol. 6. Plenum Publishing Corp., New York.
- Semler, B. L., C. W. Anderson, N. Kitamura, P. G. Rothberg, W. L. Wishart, and E. Wimmer. 1981. Poliovirus replication proteins: RNA sequence encoding P3-1b and the site of proteolytic processing. *Proc. Natl. Acad. Sci. U.S.A.* 78:3464-3468.
- Semler, B. L., R. Hanecak, C. W. Anderson, and E. Wimmer. 1981. Cleavage sites in the polypeptide precursors of poliovirus protein P2-X. *Virology* 114:589-594.
- Spackman, D. H., W. H. Stein, and S. Moore. 1958. Automatic recording apparatus for use in the chromatography of amino acids. *Anal. Chem.* 30:1190-1205.
- Summers, D. F., and J. V. Maizel. 1968. Evidence for large precursor proteins in poliovirus synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 59:966-971.
- Yogo, Y., and E. Wimmer. 1972. Polyadenylic acid at the 3'-terminus of poliovirus RNA. *Proc. Natl. Acad. Sci. U.S.A.* 69:1877-1882.