Cleavage Sites Within the Poliovirus Capsid Protein Precursors

GLENN R. LARSEN,¹ CARL W. ANDERSON,² ANDREW J. DORNER,¹ BERT L. SEMLER,¹ AND ECKARD WIMMER^{1*}

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

Received 22 June 1981/Accepted 14 August 1981

Partial amino-terminal sequence analysis was performed on radiolabeled poliovirus capsid proteins VP1, VP2, and VP3. A computer-assisted comparison of the amino acid sequences obtained with that predicted by the nucleotide sequence of the poliovirus genome allows assignment of the amino terminus of each capsid protein to a unique position within the virus polypeptide. Sequence analysis of trypsin-digested VP4, which has a blocked amino terminus, demonstrates that VP4 is encoded at or very near to the amino terminus of the polypeptide. The gene order of the capsid proteins is VP4-VP2-VP3-VP1. Cleavage of VP0 to VP4 and VP2 is shown to occur between asparagine and serine, whereas the cleavages that separate VP2/VP3 and VP3/VP1 occur between glutamine and glycine residues. This finding supports the hypothesis that the cleavage of VP0, which occurs during virion morphogenesis, is distinct from the cleavages that separate functional regions of the polypeptide.

All picornaviruses are nonenveloped, icosahedral virions consisting of a plus-strand RNA genome surrounded by approximately 60 molecules each of four virus-coded capsid proteins. These capsid proteins serve to protect the RNA genome as well as to interact with specific host-cell surface receptors during virus adsorption. The capsid proteins also elicit an antigenic response in the host animal that may limit virus infection and protect the host. Because capsid proteins play such crucial roles in the virus life cycle, their structure is of particular interest to both molecular and medical virologists.

Expression of the polio viral genome results in the synthesis of a polypeptide, NCVP00 (247,000 molecular weight [MW]), which is cleaved during translation into at least 12 major polypeptides (8, 10, 21). The initial proteolytic products, P1-1a (NCVP1a), P2-3b (NCVP3b), and P3-1b (NCVP1b), correspond to three regions of the genome, designated P1, P2, and P3, respectively (see Fig. 4). The four major capsid proteins of poliovirus, VP1 (33,500 MW), VP2 (30,000 MW), VP3 (26,400 MW), and VP4 (7,400 MW), are the final proteolytic products of the precursor P1-1a (97,000 MW). Pactamycin mapping has positioned P1-1a at the amino terminus of NCVP00 (22, 23), and a similar method was used to determine the gene order of the capsid proteins as (NH₂)-VP4-VP2-VP3-VP1 (17). In addition to the major structural proteins, virions usually contain a few copies of VP0 (37,300 MW), the immediate precursor to VP4 and VP2

(9). Cleavage of VP0 has been shown to be associated with virion morphogenesis (11).

To determine the primary structure of each capsid protein and to identify the precise sites for proteolytic processing, we have obtained partial amino acid sequence information for each of the capsid proteins of Mahoney type I poliovirus. Virion proteins were labeled in vivo with several ³H-amino acids, one at a time, or with [³⁵S]methionine. Purified virus was then directly subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis. This method did not produce doublets of viral proteins, as reported by others (24). Sequencing was accomplished by the application of the purified virion proteins to a Beckman 890C protein sequencer (4, 20). The partial amino acid sequence obtained was then compared with that predicted by the recently completed nucleotide sequence of the poliovirus genome (12).

Representative results from microsequencing experiments on VP1, VP2, and VP3 are shown in Fig. 1. For VP1, radioactivity from [³⁵S]methionine was found in residues 5 and 9 and radioactivity from [³H]glycine was found in residues 1, 3, and 19. When the translated sequence of the poliovirus genome (12) was searched for this arrangement of amino acids, only one possible encoding location was found. This analysis places the first nucleotide encoding the amino-terminal glycine of VP1 at 2,475 nucleotides from the 5' end of the genome (Fig. 2).

Sequence analysis of [³H]glycine-labeled and

FIG. 1. Partial amino-terminal amino acid sequence of poliovirus capsid proteins VP2, VP3, and VP1. Capsid proteins, labeled in vivo with a single radioactive amino acid as described in the text, were separated by preparative sodium dodecyl sulfate-polyacrylamide gel electrophoresis of purified virions. After electrophoretic elution of the polypeptides from the gel slices, individual preparations were applied directly to a Beckman 890C sequencer together with 2 mg of apomyoglobin; sequencing was accomplished with a 0.1 M Quadrol buffer program (Beckman no. 030176) as previously described (1, 20). Each panel represents a separate experiment and presents the yield of radioactivity as a function of the Edman degradation cycle number (residue). The translated nucleotide sequence corresponding to the amino terminus of each capsid protein is given in the single-letter amino acid code at the bottom of each panel pair. The amount of radioactivity applied to the sequencer was: 32,500 cpm of [^3H]serine-labeled VP2; 100,900 cpm of [^3H]leucine-labeled VP2; 70,200 cpm of [^3H]glycine-labeled VP3; 32,500 cpm of [^3H]alanine-labeled VP3; 41,900 cpm of [^{35}S]methionine-labeled VP1; 18,800 cpm of [^3H]glycine-labeled VP1. The expected amino acid composition of these capsid proteins can be deduced from the published nucleotide sequence (12).

[^3H]alanine-labeled VP3 indicated that glycine was the 1st and 9th residue, whereas alanine was found to be the 16th residue (Fig. 1). Computer analysis revealed only one genomic location compatible with this arrangement of amino acids. The amino terminus of VP3 occurs at nucleotide 1,761 (Fig. 2). This sequence location was confirmed by an analysis of [^{35}S]methionine-labeled VP3; methionine is the fifth residue of VP3 (data not shown) as predicted by the nucleotide sequence (12).

Sequence analysis of VP2 was performed in a similar manner. As shown in Fig. 1, leucine was found to be the 14th, 16th, and 18th residues, whereas the 1st, 10th, and 21st residues were found to be serine. Computer analysis places the

codon of the amino-terminal residue of VP2 at nucleotide 948 (Fig. 2). In support of this conclusion, tyrosine was found to be the 9th residue, isoleucine the 4th and 23rd residues, and glycine the 8th residue (data not shown).

Amino-terminal sequence analysis cannot be used to locate the genomic encoding location of VP4 because it has a blocked amino terminus (17, 22; A. J. Dorner, E. Wimmer, and C. W. Anderson, unpublished data). Although we have not been able to directly identify the amino terminus of VP4, we have been able to confirm the fact that it is encoded within the polyprotein to the left of VP2 and to place limits as to the position of its amino terminus. Internal amino acids within VP4 were positioned by microse-

FIG. 2. Nucleotide and codon assignments of the capsid protein junctions. Nucleotide assignments are numbered from the 5' terminus of the genome. The solid triangles represent proteolytic cleavage sites which generate the capsid proteins VP2, VP3, and VP1, respectively. The amino-terminal codons, predicted from the nucleotide sequence and confirmed by protein sequencing as described above for VP2, VP3, and VP1, are indicated by bars above and below the codon. Confirmation of predicted codons in a tryptic fragment of VP4 is also shown by bracketed bars.

sequence analysis of the mixed tryptic peptides of purified VP4. This technique has previously been used to identify peptides within bacteriophage MS2 proteins (2). Tyrosine was found to be the 2nd, 9th, 14th, and 15th residues of the peptides produced by trypsin digestion (Fig. 3). Analysis of undigested tyrosine-labeled VP4 failed to yield significant radioactivity at any residue, as expected (data not shown). Inspection of the nucleotide sequence reveals that four tyrosine residues are predicted to occur in the long open reading frame of the polyprotein to the 5' side of the amino terminus of VP2 (Fig. 2). The sequencing data agree with this predicted arrangement of tyrosine residues within the single tryptic peptide encoded by nucleotides 795 to 842 (Fig. 2). No other tyrosine-containing tryptic peptides of VP4 have been found by reverse-phase chromatographic analysis (unpublished data). Additional evidence confirming the amino acid composition of this tryptic peptide was obtained by sequence analysis of a tryptic digest of [^{3} H]isoleucine-labeled VP4 which indicated that isoleucine is the 7th and 12th residue of the tyrosine-containing tryptic peptide (data not shown).

In agreement with previous pactamycin mapping experiments (17), the data presented in Fig. 2 show that VP4 maps in the open reading frame to the 5' side of the amino terminus of VP2. We have not yet determined the exact encoding location for the amino terminus of VP4; however, a consideration of its reported molecular weight, 8,000 M_r (19), its reported lack of cysteine residues (18, 25), and the location of the first methionine codon in the long, open reading frame of poliovirus RNA leads us to suggest that the amino terminus of VP4 may be encoded at

nucleotide 744, a position 68 amino acids to the 5' side of the amino terminus of VP2 (for a discussion, see reference 12). Experiments to directly identify the amino terminus of VP4 as well as the initiation site for synthesis of the polyprotein are in progress.

We have provided direct evidence that the amino-terminal amino acid residues of VP2, VP3, and VP1 are serine, glycine, and glycine,

FIG. 3. Sequence analysis of [^{3} H]tyrosine-labeled, trypsin-digested VP4. VP4 labeled with [^{3} H]tyrosine was prepared as described in the legend to Fig. 1, and sodium dodecyl sulfate was removed from the eluted protein by three extractions with triethylamine-acetic acid-acetone (5:5:90). This VP4 preparation was digested with tolylsulfonyl phenylalanyl chloromethyl ketone-trypsin (20 μ g; Worthington Biochemical Corp.) in 0.1 M NH_4HCO_3 with 1 mg of apomyoglobin as described previously (2). The resulting peptide mixture, containing 43,900 cpm of [^{3} H]tyrosine, was applied to a Beckman 890C sequencer with 2 mg of apomyoglobin and 3 mg of Polybrene; sequencing was accomplished with the 0.1 M Quadrol buffer program. The predicted amino acid sequence of the tyrosine-containing tryptic peptide of VP4 is given at the bottom of the figure (single letter code) and in Fig. 2.

respectively (Fig. 1 and 2). The complete nucleotide sequence of poliovirus RNA (12) allows us also to predict the amino acid residue that precedes each of these amino-terminal amino acids. Hence we can predict the amino acid pair in the precursor polypeptide(s) in which cleavage occurs to generate the carboxy terminus of one protein and the amino terminus of the adjacent protein. A comparison of these cleavage sites for the capsid proteins of poliovirus (from Fig. 2) and other picornaviruses studied, mengovirus (26) and aphthovirus (3, 5), is shown in Table 1. Although no obvious consensus of cleaved amino acid pairs can be derived from Table 1, glutamine is the predominant amino acid found at the carboxy-terminal position, and glycine is the most frequent amino acid present at the amino-terminal position. For poliovirus, the VP2/VP3 cleavage and the VP3/VP1 cleavage both occur at Gln-Gly amino acid pairs. Recent reports (20, 20a) have shown that cleavage between glutamine and glycine occurs in the precursor polypeptides to produce most of the P2 and P3 region protein products in poliovirus-infected HeLa cells. These findings show that proteolytic processing of poliovirus precursors occurs predominantly at Gln-Gly amino acid pairs and suggest that a single proteinase may be responsible for these cleavages. Although the Gln-Gly sequence must be an important part of the proteolytic specificity, we assume that other recognition factors are involved since no significant degradation of cellular proteins occurs during the first 4 h of a poliovirus infection (14). In addition, a few Gln-Gly pairs are located within the poliovirus polyprotein for which we have not yet found corresponding cleavage products.

One notable exception to the processing of poliovirus proteins at Gln-Gly pairs is the proteolytic cleavage of VP0 that generates the protein products VP4 and VP2. This cleavage is predicted to occur at an Asn-Ser dipeptide in VP0 and appears to be concomitant with viral RNA association with procapsid structures, the precursors to mature virions. Procapsids have been shown to contain only three proteins, VP0, VP1, and VP3 (11). For a detailed review of picornaviral assembly, refer to a recent article

FIG. 4. Translational domains of poliovirus. Schematic illustration of the three domains of the poliovirus polyprotein, emphasizing the relationship of the capsid region products. Numbers at the termini of vertical lines give the nucleotide assignment (12) of the first nucleotide in the amino-terminal codon for the indicated product. Molecular weights ($\times 10^3$) for these products, calculated from the predicted amino acid sequences (12), are given in parentheses. Solid triangles illustrate Gln-Gly cleavage sites; the solid diamond represents the Asn-Ser cleavage site. Solid circles correspond to amino-terminal glycine residues; the open circle identifies the amino-terminal serine of VP2.

by Putnak and Phillips (16). Furthermore, in vitro processing of P1-1a using infected cytoplasmic extracts produced the polypeptide products VP0, VP1, and VP3 (13). Thus without capsid assembly, either in vitro or in vivo, there is no conversion of VP0 to VP4 and VP2. This result implies that association with viral RNA induces a conformational alteration of the procapsid, rendering VP0 sensitive to proteinase. Since the VP0 cleavage site does not conform with the predicted Gln-Gly specificity of the proteinase that produces most viral proteins, the possibility exists that another proteinase (either viral or cellular) is responsible for this cleavage. Interestingly, proteinase activity in highly purified virus preparations has been reported for poliovirus (7) and encephalomyocarditis virus (15).

Figure 4 shows the three major coding regions (P1, P2, and P3) of the viral genome and summarizes the results presented in this paper for the products encoded by the P1 region. The placement of protein P1-3a in this scheme as an intermediate precursor to VP0 and VP3 is based on the data of Butterworth (6) and our own preliminary evidence that P1-3a is blocked at its amino terminus (unpublished). The molecular weights of the capsid proteins are calculated from our sequencing data on the assumption that no carboxy-terminal trimming occurs. These values are in close agreement with previous estimates (6, 19).

We thank Lydia Fedus and Medora Hardy for excellent technical assistance. We are grateful to Paul Rothberg and

TABLE 1. Cleavage sites of picornavirus capsid proteins

Virus	Cleavage site for:		
	VP4/VP2	VP2/VP3	VP3/VP1
Poliovirus	-Asn-Ser-	-Gln-Gly-	-Gln-Gly-
Mengovirus ^a	-Ala-Asp-	-Gln-Ser-	-Gln-Gly-
Aphthovirus ^b	-Ala-Asp-	-Glu-Gly-	-Gln-Thr-

^a Reference 26.

^b References 3 and 5.

Emilio Emini for helpful discussions, to Naomi Kitamura for making the nucleotide sequence of poliovirus RNA available to us prior to publication, and to Susan Studier for preparation of the manuscript.

This work was supported by Public Health Service grants AI-15122 and CA-28146 from the National Institutes of Health, and by the United States Department of Energy. G.R.L. and A.J.D. are members of the Molecular Biology Program of the State University of New York at Stony Brook. B.L.S. is a postdoctoral fellow of the National Institutes of Health, Public Health Service grant AI-05935.

LITERATURE CITED

- Anderson, C. W., and J. B. Lewis. 1980. Amino-terminal sequence of adenovirus type 2 proteins: hexon, fiber, component IX, and early protein 1B-15K. *Virology* 104:27-41.
- Atkins, J. F., R. F. Gesteland, B. R. Reid, and C. W. Anderson. 1979. Normal tRNAs promote ribosomal frameshifting. *Cell* 18:1119-1131.
- Bachrach, H. L., J. B. Swaney, and G. F. Vande Woude. 1973. Isolation of the structural polypeptides of foot-and-mouth disease virus and analysis of their C-terminal sequences. *Virology* 52:520-528.
- Bhown, A. S., J. E. Mole, F. Hunter, and C. J. Bennett. 1980. High-sensitivity sequence determination of proteins quantitatively recovered from sodium dodecyl sulfate gels using an improved electrodialysis procedure. *Anal. Biochem.* 103:184-190.
- Boothroyd, J. C., P. E. Highfield, G. A. M. Cross, D. J. Rowlands, P. A. Lowe, F. Brown, and T. J. R. 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. 1973. A comparison of the virus-specific polypeptides of encephalomyocarditis virus, human rhinovirus-1A, and poliovirus. *Virology* 56:439-453.
- Holland, J. J., M. Doyle, J. Perrault, D. T. Kingsbury, and J. Etchison. 1972. Proteinase activity in purified animal viruses. *Biochem. Biophys. Res. Commun.* 46:634-639.
- Holland, J. J., and E. D. Klehn. 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., J. Asso, and D. Baltimore. 1970. Further evidence on the formation of poliovirus proteins. *J. Mol. Biol.* 49:657-669.
- 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.
- Jacobson, M. F., and D. Baltimore. 1968. Morphogenesis of poliovirus. I. Association of the viral RNA with coat protein. *J. Mol. Biol.* 33:369-378.
- 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.
- Korant, B. D. 1973. Cleavage of poliovirus-specific polypeptide aggregates. *J. Virol.* 12:556-563.
- Korant, B. D., N. Chow, M. Lively, and J. Powers. 1980. Proteolytic events in replication of animal viruses. *Ann. N.Y. Acad. Sci.* 343:304-318.
- Lawrence, C., and R. E. Thach. 1975. Identification of a viral protein involved in post-translational maturation of the encephalomyocarditis virus capsid precursor. *J. Virol.* 15:918-928.
- Putnak, R. J., and B. A. Phillips. 1981. Picornaviral structure and assembly. *Microbiol. Rev.* 45:287-315.
- Rekosh, D. 1972. Gene order of the poliovirus capsid proteins. *J. Virol.* 9:479-487.
- Rueckert, R. R. 1976. On the structure and morphogenesis of picornaviruses. *Compr. Virol.* 6:131-213.
- Rueckert, R. R., T. J. Matthews, O. M. Kew, M. Pal-lansch, C. McLean, and D. Omilianowski. 1979. Synthesis and processing of picornaviral polyprotein, p. 113-125. *In* R. Perez-Bercoff (ed.), *The molecular biology of picornaviruses*. Plenum Publishing Co., Ltd., 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 sites of proteolytic processing. *Proc. Natl. Acad. Sci. U.S.A.* 78:3464-3468.
- 20a. 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.
- 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.
- Summers, D. F., and J. V. Maizel. 1971. Determination of the gene sequence of poliovirus with pactamycin. *Proc. Natl. Acad. Sci. U.S.A.* 68:2852-2856.
- Taber, R., D. Rekosh, and D. Baltimore. 1971. Effect of pactamycin on synthesis of poliovirus proteins: a method for genetic mapping. *J. Virol.* 8:395-401.
- Vrijksen, R., A. Boeye, and A. D. Strosberg. 1978. Amino-terminal sequence ambiguity in three capsid polypeptides of poliovirus. *Biochem. Biophys. Res. Commun.* 85:1596-1601.
- Wouters, M., and J. Vandekerckhove. 1976. Amino acid composition of the poliovirus capsid polypeptides isolated as fluorescamine conjugates. *J. Gen. Virol.* 33:529-533.
- Ziola, B. R., and D. G. Scraba. 1976. Structure of the Mengo virion. IV. Amino- and carboxy-terminal analyses of the major capsid polypeptides. *Virology* 71:111-121.