Identification of the Initiation Site of Poliovirus Polyprotein Synthesis

ANDREW J. DORNER,^{1*} LYDIA F. DORNER,¹ GLENN R. LARSEN,^{1†} ECKARD WIMMER,¹ AND CARL W. ANDERSON²

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 28 December 1981/Accepted 16 February 1982

The complete nucleotide sequence of poliovirus RNA has a long open reading frame capable of encoding the precursor polyprotein NCVP00. The first AUG codon in this reading frame is located 743 nucleotides from the 5' end of the RNA and is preceded by eight AUG codons in all three reading frames. Because all proteins that map at the amino terminus of the polyprotein (P1-1a, VP0, and VP4) are blocked at their amino termini and previous studies of ribosome binding have been inconclusive, direct identification of the initiation site of protein synthesis was difficult. We separated and identified all of the tryptic peptides of capsid protein VP4 and correlated these peptides with the amino acid sequence predicted to follow the AUG codon at nucleotide 743. Our data indicate that VP4 begins with a blocked glycine that is encoded immediately after the AUG codon at nucleotide 743. An S1 nuclease analysis of poliovirus mRNA failed to reveal a splice in the 5' region. We concluded that synthesis of the poliovirus polyprotein is initiated at nucleotide 743, the first AUG codon in the long open reading frame.

Poliovirus-specific protein synthesis has been studied for several decades and has been used as a model of translation of eucaryotic mRNA (5, 21). The poliovirus genome is a single-stranded RNA of positive polarity (1). After infection, the poliovirus RNA directs the synthesis of a large polyprotein, NCVP00, which has been suggested to contain most, if not all, of the translatable information of the virus. This polyprotein is cleaved as a nascent chain into the primary precursor proteins P1-1a, P2-3b, and P3-1b, which correspond to three distinct classes of poliovirus proteins (9, 10, 30) (Fig. 1A). Pactamycin mapping and salt shock studies (25, 31, 33) have indicated that the three main protein classes are encoded on the RNA in the order (from the 5' end to the 3' end): P1-1a-P2-3b-P3-1b. The primary precursors are processed proteolytically to produce all of the known poliovirus-specific proteins. P1-1a, contained at the amino terminus of the polyprotein, is the precursor to the four poliovirus capsid proteins that map in the order VP4-VP2-VP3-VP1 (23). Cleavage of P1-1a occurs in distinct steps. This protein is cleaved initially to VP3, VP1, and VP0. In turn, VP0 is cleaved to VP4 and VP2. Cleavage of VP0 is intimately associated with morphogen-

esis and probably occurs upon association of vRNA with procapsid structures (24).

The genomic RNA (vRNA) has the small virus-encoded protein VPg linked to the 5' terminus (17) and is polyadenylated at the 3' terminus (37). The virus-specific RNA associated with polyribosomes (mRNA) lacks VPg and terminates with a 5'-uridylyate (6, 8, 19). The complete nucleotide sequence of the poliovirus genome has been determined recently (11, 22). This sequence analysis has revealed a long open reading frame that encompasses 89% of the viral genome and is capable of encoding the precursor polyprotein NCVP00. A partial amino acid sequence analysis of 13 viral proteins has shown that all of these proteins may be derived by processing of the polyprotein and has confirmed the genetic map of the poliovirus genome (5a, 11, 14a, 27, 28).

The nucleotide assignments used in this paper are from a "consensus" sequence for poliovirus type 1 [Mahoney] derived from the previously published sequences of Mahoney strain virus [11, 22], a sequence analysis of cloned cDNA segments derived from Mahoney virus [S. van der Werf, personal communication], and a sequence analysis of Sabin vaccine strain virus [A. Nomoto, personal communication]. This sequence is available through the MOLGEN pro-

† Present address: Genetics Institute, Boston, MA 02115.

FIG. 1. Genomic organization of poliovirus. (A) The three regions of the poliovirus polyprotein are delineated above the heavy horizontal line representing genomic RNA. A simplified processing scheme detailing the capsid protein region is presented. The numbers at the ends of the vertical lines indicate the nucleotide assignments on the genomic RNA for initiation (nucleotide 743) and termination (nucleotide 7,369) of polypeptide synthesis and for the primary protein cleavage sites (nucleotides 3,386 and 5,111). The numbers in parentheses are the molecular weights ($\times 10^{-3}$) calculated from the predicted amino acid sequence. The solid circles indicate experimentally determined blocked amino termini of processed proteins. The unprocessed polyprotein (NCVPO0) was not examined directly; the amino terminus of this protein is indicated by an open circle. poly(A), Polyadenylic acid. (B) Positions of AUG codons and termination codons within the 5'-terminal 1,000 nucleotides of the poliovirus genome. The numbers on the left indicate the three reading frames beginning with the first, second, or third nucleotide of the RNA. The vertical lines represent termination codons. The solid boxes represent AUG codons, and the open boxes indicate open translation regions following an AUG codon. The long open reading frame (nucleotides 743 through 7,369) present in the second reading frame is labeled NCVPO0. The numbers below the vertical arrows indicate distances from the 5' end (in nucleotides). The nucleotide assignments are from a "consensus" sequence (see text).

ject of the Stanford University SUMEX-AIM facility. Differences between previously published sequences and the "consensus" sequence do not affect the conclusions reached in this paper.

Unexpectedly, an analysis of the poliovirus RNA sequence showed that the first AUG codon in the long open reading frame is located 743 nucleotides from the 5' end of the RNA and is preceded by eight other AUG codons which occur in all three reading frames. Translation initiated at the 5'-proximal AUG codon at nucleotide 185 could produce a protein of 4,720 daltons; the eighth AUG codon, located at nucleotide 586, is followed by an open reading frame of 64 codons, which overlaps the beginning of the long open reading frame. The six other AUG codons are followed shortly by in-phase termi-

nation codons (Fig. 1B). The existence of these potential initiation codons has raised questions concerning the exact location of the initiation site for synthesis of the poliovirus polyprotein.

Nucleotide sequence analysis of poliovirus RNA alone did not identify the initiation site of polyprotein synthesis. The possibility of sequencing errors or of an RNA splicing event necessitated an analysis of viral mRNA and protein sequences. Because all of the proteins which map at the amino terminus of the poliovirus polyprotein (P1-1a, VP0, VP4) carry blocked amino termini, direct protein sequencing has not been possible. Ribosome binding experiments have also failed to produce the initiation site sequence (Dorner and Wimmer, unpublished data; E. Ehrenfeld, personal communication). We located the initiation site for synthesis of the

poliovirus polypeptide by mapping the tryptic peptides of VP4 within the genomic 5' coding sequence. Our strategy involved identifying all of the tryptic peptides of VP4 and determining the composition of the amino-terminal tryptic peptide. We found no evidence of a splice within the 5' region of poliovirus mRNA. Together, the protein and RNA analyses demonstrate that poliovirus polypeptide synthesis is initiated at the first AUG codon in the long open reading frame of the poliovirus genomic sequence.

MATERIALS AND METHODS

Preparation of radiolabeled poliovirus proteins. Suspension cultures of HeLa S3 cells (5×10^6 cells per ml) were infected with poliovirus type 1 (Mahoney) at a multiplicity of 50 PFU/cell. Infected cultures were labeled with radioactive amino acids (Amersham Corp.; ICN) at concentrations of 0.04 to 0.06 mCi/ml for 5 h beginning 2 h postinfection. Harvested cells were lysed by three rounds of freeze-thawing in reticulocyte standard buffer containing Mg^{2+} (0.01 M NaCl, 0.01 M Tris, pH 7.4, 1.5 mM $MgCl_2$), and the cell nuclei were pelleted. The supernatant containing poliovirus virions was made 2 mM in EDTA and 0.1% in sodium dodecyl sulfate (SDS) and applied to a 15% to 30% sucrose gradient (0.1 M NaCl, 0.01 M Tris, pH 7.4, 1 mM EDTA, 0.5% SDS). Centrifugation was at $16,000 \times g$ for 16 h. Gradient-purified virus was pelleted, suspended in protein gel sample buffer, and subjected to preparative SDS-polyacrylamide gel electrophoresis (5a). The band corresponding to VP4 was excised from the gel and eluted electrophoretically (27). P1-1a and VP0 were purified from infected cell extracts as described previously (27).

Trypsin digestion of VP4. Apomyoglobin (1 mg) was added as a carrier to the gel-purified VP4. SDS removal was accomplished by three 30-min extractions with ice-cold acetone-acetic acid-triethylamine (90:5:5), followed by two washes with ice-cold acetone (7). After drying, the protein precipitate was incubated on ice for 1.5 h with 0.5 ml of performic acid (19:1 mixture of 88% formic acid and 30% hydrogen peroxide kept at room temperature for 2 h). The oxidized protein was lyophilized three times. Then the protein was suspended in 0.4 ml of 1% ammonium bicarbonate (pH 8.5) and incubated with 10 μ g of tolylsulfonyl phenylalanyl chloromethyl ketone-treated trypsin (Worthington Diagnostics) for 2 h at 37°C. An additional 20 μ g of trypsin was added, and the incubation was continued for 2 h. Then 10 μ g of trypsin was added again, and the incubation was continued for a final 2 h (total incubation time, 6 h). The reaction mixture was then lyophilized. The digested protein was dissolved in 250 μ l of 20% formic acid for high-performance liquid chromatography or small volumes of 15% NH_4OH or 20% pyridine for application to cellulose thin-layer plates.

Two-dimensional tryptic peptide mapping. Separation in the first dimension was by ascending chromatography on cellulose-coated thin-layer plates (20 by 20 cm; EM Reagents) in *n*-butanol-pyridine-acetic acid-water (30:20:6:24) until the solvent front was 2 cm from the top of the plate. Plates were dried overnight at room temperature. Electrophoresis in the second dimension was for 30 to 40 min at 1,000 V (40 to 50 mA) in pH 3.5 buffer containing acetic acid,

pyridine, and water (10:1:89). Dried plates were sprayed with En³Hance (New England Nuclear Corp.) and exposed at -70°C to Kodak XAR-5 film.

Amino acid analysis. A 100- μ g portion of electrophoretically purified VP4 was dialyzed extensively against distilled water to remove SDS and other low-molecular-weight components. The sample was divided into four equal portions, three of which were hydrolyzed in 0.3 ml of 6 N HCl at 115°C for 24, 48, or 72 h in evacuated, sealed borosilicate tubes. After hydrolysis, each sample was analyzed for amino acid content with a single-column analyzer. A standard amino acid mixture was analyzed before and after the set of experimental samples, and this standard was used to calculate the amount of each amino acid in each experimental sample. To convert the data to residue equivalents, the values for six stable amino acids (alanine, asparagine-aspartic acid, glutamine-glutamic acid, leucine, phenylalanine, valine) were added together and divided by the total number of residues of these six amino acids predicted by the sequence between nucleotides 746 and 949. Division by these factors (24 h, 1.725; 48 h, 1.688; 72 h, 1.685) gave the number of residues of each amino acid present in VP4. The values for threonine and serine for each hydrolysis time were plotted and extrapolated to zero time in order to correct for the destruction of these residues. No other amino acid showed a significant trend with hydrolysis time, and so the values for all three hydrolysis times were averaged.

High-performance, reverse-phase (C18) liquid chromatography. A mixture of radiolabeled tryptic peptides of VP4 dissolved in 250 μ l of 20% formic acid was applied to a LiChrosorb RP-18, 10 μ m column, 250 by 4.6 mm; EM Reagents). The column was developed with a discontinuous linear gradient made from 20 mM sodium phosphate (pH 5.8) (buffer A) and 100% acetonitrile (buffer B). The flow rate was 2.0 ml/min, and 1.0-ml fractions were collected. A portion of each fraction was counted with Aquasol II (New England Nuclear Corp.) by using a Searle Mark III scintillation counter.

Amino acid sequence analysis. Sequence analyses of purified radiolabeled proteins or tryptic peptides were performed essentially as described previously (27).

S1 nuclease analysis. [³H]uridine-labeled poliovirus vRNA and mRNA were purified essentially as described previously (15, 19). pPV1-366, a pBR322 derivative containing a 3.05-kilobase insertion completely spanning the 5' end of the poliovirus genome, was obtained from S. van der Werf (34). *EcoRI* (New England Biolabs) cuts this plasmid at a single site outside the poliovirus-specific sequences and was used to linearize pPV1-366. Hybridization of radiolabeled poliovirus vRNA and mRNA with linearized pPV1-366 was performed in 10- μ l reaction mixtures containing 80% formamide, 0.4 M NaCl, 0.01 M PIPES [piperazine-*N,N'*-bis(2-ethanesulfonic acid)] (pH 6.4), and 0.001 M EDTA (2). A poliovirus-specific DNA-RNA single-stranded sequence ratio of 3:1 was used for hybridization. The samples were heated for 10 min at 60°C and then incubated for 2.5 h at 55.5°C. The hybridization reaction was terminated by adding 8 to 10 volumes of ice-cold 0.2 M NaCl-0.05 M sodium acetate (pH 4.5)-0.001 M zinc sulfate. S1 nuclease (Bethesda Research Laboratories) was added (concentration range, 50 to 1,000 U/ml), and the samples were

incubated for 45 min at 37°C. Then 10 µg of yeast tRNA and 2.5 volumes of chilled 100% ethanol were added, and the nucleic acids were precipitated at -20°C overnight. After centrifugation, the dried nucleic acid pellet was dissolved in water, adjusted to 10% glycerol, 0.01% bromophenol blue, 0.04 M Tris-acetate (pH 8.1), and 0.002 M EDTA, and subjected to electrophoresis in 0.7% agarose gels in 0.04 M Tris-acetate (pH 8.1)-0.002 M EDTA. The agarose gels were treated with En³Hance (New England Nuclear Corp.), dried, and exposed at -70°C to Kodak XAR-5 film.

RESULTS

P1-1a, VP0, and VP4 have blocked amino termini. Pactamycin and salt shock mapping studies, molecular weight considerations, and the known genomic sequence of poliovirus suggest that NCVP0, P1-1a, VP0, and VP4 share the same amino-terminal sequence. To prove this hypothesis, we attempted to determine the amino-terminal sequences of P1-1a, VP0, and VP4 by direct radiochemical sequence analysis.

VP4 labeled with [³H]tyrosine, [³H]valine, or [³H]glutamine was purified as described above and subjected to automated Edman degradation. No significant yield of radioactivity was obtained in any of the first 21 degradation cycles (data not shown). An analysis of the carrier apomyoglobin from the same experiments indicated that the sequencer was functioning with a repetitive yield of over 95%. The amino acid sequence shown in Table 1 indicates that if VP4 has the amino-terminal sequence predicted for initiation of polypeptide synthesis at the AUG codon at nucleotide 743, tyrosine, valine, and glutamine residues would be present within the first 20 amino acids. Proteins purified and eluted by the method which we used for VP4 have

rarely failed to yield at least 30% of the theoretically expected radioactivity unless the amino terminus was blocked (Anderson, in *Genetic Engineering Principles and Methods*, vol. 4, in press). As described previously (14a), automated Edman degradation of a mixture of [³H]tyrosine-labeled tryptic peptides of VP4 gave a pattern of tyrosine residues consistent with the predicted tryptic peptide 3 (Table 1). A sequence analysis of a high-performance liquid chromatography-purified, [³H]histidine-labeled tryptic peptide indicated that there was a histidine residue at position four, which was consistent with the tryptic peptide 2 of VP4 (data not shown). This data established the correct reading frame and ruled out the possibility of extensive amino-terminal processing of VP4. We concluded that VP4 must have a blocked amino terminus.

An analysis of [³H]valine-labeled VP0 and [³H]lysine-labeled P1-1a isolated from poliovirus-infected cells also failed to yield significant radioactivity in any residue during automated Edman degradation (data not shown). These results indicate that VP0 and P1-1a also have blocked amino termini. We have failed in attempts to synthesize unblocked P1-1a in vitro by preventing protein acetylation (20). The reason for this is unknown; our data do not exclude the possibility that acetylation of the nascent polypeptide is the blocking mechanism.

Reverse-phase chromatography of VP4 tryptic peptides and identification of the carboxy-terminal peptide. Table 1 shows the expected tryptic peptides of VP4, as predicted from the RNA sequence for a protein mapping to the left of VP2 and beginning at the first AUG codon (nucleotide 743) in the long open reading frame of poliovirus vRNA. This sequence contains 68

TABLE 1. Predicted tryptic peptides of VP4

Peptide	Sequence ^a	Labeling pattern of tryptic peptides ^b								
		Gln	Ser	Ala	Val	Gly	Lys	Tyr	Ile	Met
1	B-Gly-Ala-Gln-Val-Ser-Ser-Gln-Lys	X	X	X	X	X	X			
2	Val-Gly-Ala-His-Glu-Asn-Ser-Asn-Arg ^c		X	X	X	X				
3	Ala-Tyr-Gly-Gly-Ser-Thr-Ile-Asn-Tyr-Thr-Thr-Ile-Asn-Tyr-Tyr-Arg ^d		X	X		X		X	X	
4	Asp-Ser-Ala-Ser-Asn-Ala-Ala-Ser-Lys			X	X		X			
5	Gln-Asp-Phe-Ser-Gln-Asp-Pro-Ser-Lys	X	X				X			
6	Phe-Thr-Glu-Pro-Ile-Lys						X		X	
7	Asp-Val-Leu-Ile-Lys				X		X		X	
8	Thr-Ala-Pro-Met-Leu-Asn ^e			X						X

^a Tryptic peptide sequences were predicted from the RNA sequence between nucleotide 746 (Gly) and nucleotide 949 (Asn). B, Blocked amino terminus.

^b The amino acids listed were used to radiolabel VP4. X, Presence of the amino acid in the tryptic peptide sequence.

^c The histidine residue was confirmed by automated Edman degradation.

^d The tyrosine residues were confirmed by automated Edman degradation.

^e The alanine and methionine residues were confirmed by automated Edman degradation.

FIG. 2. High-performance, reverse-phase liquid chromatography of the tryptic peptides of VP4 labeled with [^{35}S]methionine (A) or [^3H]alanine (B). The column was developed with a discontinuous linear 5 to 95% gradient of acetonitrile as depicted (A) and as described in the text. The amounts of radioactivity applied and recovered were as follows: [^{35}S]methionine, 153,000 cpm applied, 80% recovered in peak M1; [^3H]alanine, 37,100 cpm applied, 66% recovered in peaks A1 through A4.

amino acids after the initiator methionine and should produce eight peptide fragments upon complete trypsin digestion. Table 1 shows that the carboxy-terminal peptide is predicted to have methionine as its fourth amino acid and alanine as its second amino acid.

After trypsin digestion, a mixture of radiolabeled peptides of VP4 was fractionated by reverse-phase liquid chromatography, as described above. Figure 2 shows chromatograms exhibiting the profiles of peptides obtained from [^{35}S]methionine- or [^3H]alanine-labeled VP4. Only a single methionine-containing peak (peptide M1) was resolved, which contained 80% of the radioactivity applied to the column. Four alanine-containing peptides were resolved, one of which (peptide A2) chromatographed at the same position as the methionine-containing peptide. Peptides M1 and A2 were subjected to automated Edman degradation (Fig. 3). An ami-

no-terminal sequence analysis indicated that methionine was the fourth residue and alanine was the second residue of this peptide. These data identified the carboxy-terminal peptide of VP4. Recently, asparagine has been found to be the carboxy-terminal amino acid of VP4 (5a), a result that is also in agreement with the predicted carboxy-terminal sequence. Together, these analyses demonstrate that VP2 and VP4 are contiguous in the polypeptide. Since 80% of the methionine label in VP4 was recovered in the carboxy-terminal peptide, the initiating methionine must have been removed from VP4.

The results of a sequence analysis of additional radiolabeled tryptic peptides of VP4 fractionated by reverse-phase chromatography suggested that the amino-terminal peptide predicted for VP4 was present in the flow-through peak along with several other peptides (data not shown). A limited analysis of tryptic peptides by cation-

FIG. 3. Partial sequence analysis of the VP4 tryptic peptide corresponding to the carboxy terminus. The single [³⁵S]methionine-labeled peptide (M1) and the corresponding [³H]alanine-labeled peptide (A2) (see Fig. 2) were subjected to automated Edman degradation. Sequencing was carried out for 11 cycles after an initial cycle (cycle 0) in which the addition of the cleavage reagent (heptafluorobutyric acid) was omitted. The total radioactivity released in the amino acid fraction is plotted as a function of the number of

exchange chromatography also indicated that the amino-terminal peptide was likely to be found in the flow-through peak. Consequently, alternative methods of peptide analysis were used.

Amino acid analysis of VP4. Before proceeding with additional tryptic peptide analyses, we determined the amino acid composition of VP4 and compared it with the composition predicted from the genomic sequence. A previously determined amino acid composition for VP4 (36) agrees only moderately well with the predicted composition. Our analysis (Table 2) indicated that VP4 should contain one methionine residue, one histidine residue, and no cysteine residue, in agreement with the RNA sequence and our tryptic peptide analysis. The results of the amino acid analysis for most other residues agreed with the composition predicted from the 68-amino acid sequence that begins with the glycine following the methionine residue at nucleotide 743.

The amount of glycine determined by our analysis was 1.8 residues more than the predicted value. However, glycine is frequently found as a contaminant in gel-purified protein and can also be derived from nonprotein contaminants during acid hydrolysis. The values found for alanine, glutamine-glutamic acid, and valine agree well with the predicted values. These amino acids are predicted to be in the amino-terminal tryptic peptide. Serine, also predicted to be in the amino-terminal tryptic peptide, was found at a lower level than predicted. Serine is unstable during acid hydrolysis, and extrapolation of its value from the three hydrolysis times may not have corrected for its loss. Thus, our amino acid analysis of VP4 supported the sequence shown in Table 1 but was not sufficient to establish the composition of the amino-terminal peptide.

Two-dimensional thin-layer analysis of the amino terminus of VP4. To determine the composition of the amino terminus of VP4, radiolabeled VP4 was digested with trypsin, and the resulting peptide mixture was analyzed by two-dimensional thin-layer techniques, as described above. Table 1 shows the labeling pattern for each predicted tryptic peptide with the radioactive amino acids used in this study. Our choice of labels allowed the identification of the chromatographic positions of all eight predicted tryptic peptides of VP4. Figure 4 shows the chro-

Edman degradation cycles. A total of 2,300 cpm of [³H]alanine-labeled peptide (A) and 3,800 cpm of [³⁵S]methionine-labeled peptide (B) were applied to the sequencer. The predicted sequence of the carboxy-terminal tryptic peptide is given at the top of the figure in the single-letter amino acid code.

TABLE 2. Amino acid composition of VP4

Amino acid	Predicted no. of residues ^a	Experimental no. of residues ^b
Ala	7	6.4
Arg	2	2.1
Asx ^c	10	9.1
Cys	0	<0.2
Glx ^d	6	6.4
Gly	4	5.8
His	1	1.1
Ile	4	3.7
Leu	2	2.8
Lys	5	4.4
Met	1	1.1
Phe	2	2.0
Pro	3	3.8
Ser	9	7.0
Thr	5	4.3
Trp	0	ND ^e
Tyr	4	2.4
Val	3	3.4

^a The predicted numbers of residues were calculated from the poliovirus nucleotide sequence between nucleotides 746 and 949.

^b Experimental values were determined as described in the text.

^c Asx, Asparagine-aspartic acid.

^d Glx, Glutamine-glutamic acid.

^e ND, Not determined.

matograms obtained from tryptic digests of [³⁵S]methionine-, [³H]alanine-, [³H]serine-, and [³H]glutamine-labeled VP4 preparations. The positions identified for each of the eight peptides are numbered according to their order, as indicated in the sequence given in Table 1.

Position 1 in Fig. 4 is the position of a peptide containing glycine, alanine, glutamine, valine, serine, and lysine but not tyrosine or isoleucine. This peptide had a high mobility during chromatography but was almost neutral during electrophoresis at pH 3.5. This behavior was expected for a peptide with the composition of tryptic peptide 1 which is blocked at its amino terminus. An analysis of the [³⁵S]methionine-labeled digest revealed a single tryptic peptide, identified as the carboxy-terminal peptide, which was found near the middle of the chromatogram away from the position of peptide 1. We concluded that peptide 1 is the blocked amino-terminal peptide of VP4 and that this peptide has the following amino acid sequence: block-Gly-Ala-Gln-Val-Ser-Ser-Gln-Lys. This sequence is encoded immediately after the AUG codon at nucleotide 743. Therefore, this AUG codon must be the initiation codon for synthesis of the poliovirus polyprotein which contains VP4 at its amino terminus.

This conclusion is valid only if position 1 on

the thin-layer chromatograms contained a single tryptic peptide. We consider the migration of more than one peptide at this position unlikely because the positions of all of the predicted tryptic peptides of VP4 have been identified and are consistent with their expected compositions. None of these positions overlaps with the position of the amino-terminal peptide. Only the single tyrosine-containing tryptic peptide did not consistently resolve well in this system, perhaps due to damage to the tyrosine residues during oxidation. The yield of this tryptic peptide varied from preparation to preparation and was low on the chromatograms of [³H]alanine- and [³H]serine-labeled VP4 preparations. However, analyses of [³H]isoleucine- and [³H]tyrosine-labeled VP4 preparations (data not shown) did allow identification of the position of tryptic peptide 3 (Fig. 4).

S1 nuclease analysis of poliovirus RNA. Eight AUG codons precede the initiation codon at nucleotide 743 (Fig. 1B). It is conceivable that there is a processing event which moves the initiation codon closer to the 5' end of the RNA or brings one of the small open reading frames in phase with the long open reading frame. Such a splice could produce an mRNA capable of directing the synthesis of a leader polypeptide that precedes the capsid proteins. Therefore, we analyzed whether the RNA associated with polyribosomes differed from the genomic RNA in its 5'-terminal sequences. Previous analyses of these two RNAs have shown that they both have a sedimentation coefficient of 35S (32), produce identical fingerprints of RNase T₁-resistant oligonucleotides as determined by two-dimensional gel electrophoresis (19), and have identical 5'-terminal nonanucleotides (18). As a specific probe for the 5' genomic RNA sequence, we used a pBR322 derivative (pPV1-366) which contains a cDNA insertion corresponding to nucleotides 1 to 3,050 of the poliovirus vRNA (34). Our strategy was to compare the S1 nuclease-treated hybrids produced when vRNA and mRNA were hybridized to pPV1-366. Electrophoresis on neutral agarose gels should have resolved S1 nuclease-resistant duplex molecules equal in length to the 5' RNA sequences. If a splice removed a portion of the genomic 5'-terminal sequence, then the pattern of S1 nuclease-resistant bands should not have been the same for both vRNA and mRNA.

[³H]uridine-labeled RNAs purified from poliovirions (vRNA) and from the polyribosomes of poliovirus-infected HeLa cells (mRNA) were hybridized to linearized pPV1-366. Hybrid molecules were digested with three concentrations of S1 nuclease and analyzed by electrophoresis on neutral 0.7% agarose gels. Digestion at the higher S1 nuclease concentrations produced a pre-

FIG. 4. Two-dimensional mapping of the tryptic peptides of VP4. Radiolabeled VP4 was digested with trypsin as described in the text, and the digestion products were spotted at the origin (arrow). First-dimension ascending chromatography was from bottom to top. Second-dimension electrophoresis was in the horizontal direction; the cathode was at the right. The numbers to the left of the spots correspond to the peptide numbers in Table 1. (A) [^3H]alanine-labeled VP4. The broken circles indicate the positions of tryptic peptides not labeled with [^3H]alanine. (B) [^{35}S]methionine-labeled VP4. (C) [^3H]serine-labeled VP4. (D) [^3H]glutamine-labeled VP4.

dominant band of approximately 3,000 base pairs when vRNA (Fig. 5, lanes 10 and 11) or mRNA (Fig. 5, lanes 7 and 8) was hybridized with pPV1-366. The size of this fragment is consistent with the length of the poliovirus-specific insertion (3,050 base pairs). Minor S1 nuclease-resistant bands of 2,300, 1,800, and 800 base pairs were also present in these lanes. These bands may have been due to cutting at adenine-thymine-rich regions of the duplex molecules at high S1 nuclease concentrations, a phenomenon reported by other workers (16).

The positions of several adenine-thymine-rich regions in poliovirus vRNA are consistent with this explanation. Digestion at the lowest S1 nuclease concentration (Fig. 5, lanes 6 and 9) did not produce this pattern of minor bands, and the predominant S1 nuclease-resistant band was slightly larger and more diffuse than that observed at higher S1 nuclease concentrations. This pattern is consistent with less complete digestion. Hybridization of mRNA or vRNA with pBR322 followed by S1 nuclease digestion did not produce any S1 nuclease-resistant bands

FIG. 5. S1 nuclease analysis of [^3H]uridine-labeled poliovirus RNA hybridized to a 5'-specific DNA probe (pPV1-366). The products of hybridization were digested with varying concentrations of S1 nuclease and were resolved by neutral agarose gel electrophoresis as described in the text. The autoradiogram of the gel is shown. Lanes 3 and 14 contained *Hind*III restriction digests of ^{32}P -labeled λ DNA, which were used as chain length markers. The numbers on the right indicate the lengths (in kilobase pairs) of the λ DNA fragments. Lane 1, mRNA; lane 2, vRNA; lane 4, mRNA hybridized to pPV1-366; lane 5, mRNA subjected to hybridization conditions and digested with 50 U of S1 nuclease per ml; lanes 6 through 8, mRNA hybridized to pPV1-366 and digested with 50, 500, and 1,000 U of S1 nuclease per ml, respectively; lanes 9 through 11, vRNA hybridized to pPV1-366 and digested with 50, 500, and 1,000 U of S1 nuclease per ml, respectively; lane 12, vRNA subjected to hybridization conditions and digested with 50 U of S1 nuclease per ml; lane 13, vRNA hybridized to pPV1-366; lane 15, mRNA hybridized to pBR322 and digested with 50 U of S1 nuclease per ml; lane 16, vRNA hybridized to pBR322 and digested with 50 U of S1 nuclease per ml.

(Fig. 5, lanes 15 and 16), nor did digestion of mRNA or vRNA alone (Fig. 5, lanes 5 and 12). Thus, our analysis detected no difference between S1 nuclease-treated vRNA-DNA hybrids and S1 nuclease-treated mRNA-DNA hybrids. We estimate that a size difference of 50 base pairs would have been detected by our methods. A small splice that produced a denaturation loop substantially resistant to S1 nuclease digestion might not have been observed. However, this possibility seems unlikely since sufficient S1 nuclease was used to cut apparently duplexed molecules at regions rich in adenine and thymine. A similar S1 nuclease analysis in which [^3H]thymidine-labeled pPV1-366 and alkaline agarose gel electrophoresis were used also de-

tected no difference between vRNA and mRNA (data not shown). Inspection of the 5'-terminal 1,000 nucleotides of poliovirus vRNA failed to reveal good consensus donor or acceptor splice sites (29). We concluded that mRNA and vRNA do not differ except for the presence or absence of VPg linked to the 5' end.

DISCUSSION

We identified the AUG codon at nucleotide 743, the first AUG codon in the long open reading frame of poliovirus RNA, as the initiation codon for the synthesis of polyprotein NCVP00. This was accomplished by first locating the encoding site for the amino terminus of VP4, the protein encoded near the 5' end of

poliovirus RNA and amino terminal with respect to all known poliovirus proteins derived from the polyprotein. Second, we showed that the polyribosome-associated mRNA has a 5'-terminal sequence identical to that of the vRNA. This precludes the synthesis of a leader polypeptide, a situation in which VP4 would not be located at the amino terminus of the polyprotein. Such a peptide appears to be synthesized during aphthovirus RNA translation (26).

The analysis of VP4 described in this paper identified all of the tryptic peptides of VP4 predicted by the RNA sequence to follow the AUG codon at nucleotide 743. The labeling pattern of the amino-terminal tryptic peptide showed that it contained glycine but not methionine. A glycine residue is predicted to follow immediately the initiator methionine and is the only glycine contained within the amino-terminal tryptic peptide. Our data indicate that the encoding sequence of VP4 begins immediately after the AUG codon at nucleotide 743 and that protein synthesis is followed by removal of the initiator methionine residue and addition of a blocking group to the amino-terminal glycine residue. Removal of initiating methionine residues which are followed by glycine and subsequent acetylation of the amino-terminal glycine is expected (35). Apparently no other amino-terminal processing occurs, and the amino terminus of the poliovirus polyprotein is conserved through the precursor proteins P1-1a and VP0 and VP4.

The initiation codon at nucleotide 743 is not the AUG codon nearest the 5' end of genomic RNA. Eight other AUG codons in all three reading frames are within the 5'-proximal region. We have not detected a spliced polyribosome-associated mRNA from which these sequences have been removed. Our analysis indicates that poliovirus polyprotein synthesis is not initiated at the 5'-proximal AUG codon, and thus poliovirus mRNA represents another exception to the original ribosome scanning model. Recently, a modification of the scanning model (14) has been proposed, which suggests that the sequences flanking the functional initiation codon facilitate the recognition of the correct codon by the ribosomes. Among the nine AUG codons present within the 5'-proximal RNA sequence, only the codon at nucleotide 743 has the favored initiation site sequence CAXX-AUGG. It is possible that if the 40S ribosomal subunits migrate along the 5'-terminal untranslated region of the poliovirus RNA, they bypass the upstream AUG codons due to unfavorable flanking sequences and form a functional initiation complex only at the correct initiation codon. Classical ribosome binding studies with poliovirus RNA in our laboratory and other

laboratories (E. Ehrenfeld, personal communication) have failed to determine a unique initiation site sequence. This phenomenon could be explained if 40S ribosomal subunits hesitate in their migration at the upstream AUG codons and so protect these sequences during ribosome binding experiments.

Poliovirus mRNA is uncapped, and the secondary structure of the RNA may play an important role in the recognition of the correct initiation codon by ribosomes. Denaturation mapping studies of poliovirus type 1 replicative form (RF) molecules show a distinct denaturation region located approximately 10% of the genome length from the 5' end of the positive RNA strand (4). This denatured region appears to map at an adenine-uridine-rich segment which stretches from nucleotide 687 up to and including the initiation codon at nucleotide 743. This is the longest adenine-uridine-rich region within the 5'-terminal 3,000 nucleotides of poliovirus RNA. Consequently, this region may not participate in the formation of stable secondary structures within the single-stranded RNA, thus leaving the initiation codon at nucleotide 743 exposed.

In vitro labeling studies with *N*-formyl [³⁵S]methionine have provided evidence that poliovirus RNA utilizes two initiation sites for protein synthesis in a cell-free system (3, 12). A high-molecular-weight *N*-formyl [³⁵S]methionine-labeled product apparently corresponds to P1-1a, the precursor protein from which VP4 is derived. A low-molecular-weight polypeptide (*M_r*, 5,000 to 10,000) that is synthesized at high magnesium concentrations is also labeled with *N*-formyl [³⁵S]methionine. In vitro, under conditions of low magnesium concentration, 40S ribosomal subunits have been observed to migrate more extensively along the RNA than at high concentrations (13). Accordingly, at high magnesium concentrations initiation may occur more readily at one of the AUG codons which precede the initiation codon for polyprotein synthesis. This would result in the synthesis of a small *N*-formyl [³⁵S]methionine-labeled polypeptide. For example, initiation at the 5'-proximal AUG codon would produce a 4,720-dalton protein, and initiation at the eighth AUG codon from the 5' end would produce a 7,180-dalton protein.

Speculation concerning the in vitro *N*-formyl [³⁵S]methionine labeling studies underscores the potential of the 5'-proximal region of poliovirus RNA to code for the synthesis of several small proteins independent of the initiation of polyprotein synthesis. The function of the small open reading frames which precede the AUG codon at nucleotide 743 is unknown at present. One obvious possibility is that these regions are translated, perhaps under selective conditions in vivo,

and that the polypeptides have escaped detection. An intriguing idea is that interactions of the 40S ribosomal subunits with the AUG codons preceding the polyprotein initiation codon may function as an attenuation mechanism to balance protein and RNA syntheses during the initial stages of poliovirus infection.

Alternatively, the long 5'-proximal sequence of plus-strand RNA may be essential only for RNA synthesis or for packaging of the vRNA during morphogenesis. In this case, the protein-encoding potential of this sequence is not the primary factor in its evolution, and the long 5'-proximal region may interfere with efficient protein synthesis. Removal of these sequences may make the poliovirus RNA a better messenger but render it noninfectious. However, removal of 5' genomic sequences by splicing *in vivo* may not be possible because poliovirus replicates in the cytoplasm and so may not have access to the nuclear splicing mechanisms. Resolution of these and other speculations awaits the development of defined poliovirus mutants by recombinant DNA methodology.

ACKNOWLEDGMENTS

We thank P. G. Rothberg and B. Semler for many helpful discussions, M. Hardy and N. Alonzo for excellent technical assistance, J. Demian, J. Schirmer, and A. Newlin for the preparation of artwork, K. Thompson for assistance in computing data, and S. Studier for typing the manuscript. Computer services were provided by the Stanford MOLGEN project of the National Institutes of Health SUMEX-AIM facility and by Brookhaven National Laboratory.

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.

LITERATURE CITED

- Baltimore, D. 1971. Expression of animal virus genomes. *Bacteriol. Rev.* 35:235-241.
- Berk, A. J., and P. A. Sharp. 1978. Spliced early mRNAs of simian virus 40. *Proc. Natl. Acad. Sci. U.S.A.* 75:1274-1278.
- Celma, J. L., and E. Ehrenfeld. 1975. Translation of poliovirus RNA *in vitro*: detection of two different initiation sites. *J. Mol. Biol.* 98:761-780.
- Cumakov, I. M., G. Y. Lipskaya, and V. I. Agol. 1979. Comparative studies on the genomes of some picornaviruses: denaturation mapping of replicative form RNA and electron microscopy of heteroduplex RNA. *Virology* 92:259-270.
- Darnell, J. E., and L. Levintow. 1960. Poliovirus protein: source of amino acids and time course of synthesis. *J. Biol. Chem.* 235:74-77.
- Emmini, E. A., M. Elzinga, and E. Wimmer. 1982. Carboxy-terminal analysis of poliovirus proteins: the termination of poliovirus RNA translation and the location of unique poliovirus polyprotein cleavage sites. *J. Virol.* 42:194-199.
- Fernandez-Munoz, R., and J. E. Darnell. 1976. Structural difference between the 5' termini of viral and cellular mRNA in poliovirus-infected cells: possible basis for the inhibition of host protein synthesis. *J. Virol.* 18:719-726.
- Henderson, L. E., S. Oroszlan, and W. Konigsberg. 1979. A micromethod for complete removal of dodecyl sulfate from proteins by ion-pair extraction. *Anal. Biochem.* 93:153-157.
- Hewlett, M. J., J. K. Rose, and D. Baltimore. 1976. 5'-Terminal structure of poliovirus polyribosomal RNA is pUp. *Proc. Natl. Acad. Sci. U.S.A.* 73:327-330.
- 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., B. L. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emmini, 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.
- Knauert, F., and E. Ehrenfeld. 1979. Translation of poliovirus RNA *in vitro*: studies on n-formylmethionine-labeled polypeptides initiated in cell-free extracts prepared from poliovirus infected HeLa cells. *Virology* 93:537-546.
- Kozak, M. 1979. Migration of 40S ribosomal subunits on a messenger RNA when initiation is perturbed by lowering magnesium or adding drugs. *J. Biol. Chem.* 254:4731-4738.
- Kozak, M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* 9:5233-5252.
- Larsen, G. R., C. W. Anderson, A. J. Dorner, B. L. Semler, and E. Wimmer. 1982. Cleavage sites within the poliovirus capsid protein precursors. *J. Virol.* 41:340-344.
- Larsen, G. R., B. L. Semler, and E. Wimmer. 1981. Stable hairpin structure within the 5'-terminal 85 nucleotides of poliovirus RNA. *J. Virol.* 37:328-335.
- Miller, K. G., and B. Sollner-Webb. 1981. Transcription of mouse rRNA genes by RNA polymerase 1: *in vitro* and *in vivo* initiation and processing sites. *Cell* 27:165-174.
- Nomoto, A., B. Detjen, R. Pozzatti, and E. Wimmer. 1977. The location of the polio genome protein in viral RNAs and its implication for RNA synthesis. *Nature (London)* 268:208-213.
- Nomoto, A., N. Kitamura, F. Golini, and E. Wimmer. 1977. The 5'-terminal structures of poliovirus RNA and poliovirus mRNA differ only in the genome-linked protein VPg. *Proc. Natl. Acad. Sci. U.S.A.* 74:5345-5349.
- Nomoto, A., Y. F. Lee, and E. Wimmer. 1976. The 5' end of poliovirus mRNA is not capped with m⁷G(5')ppp(5')Np. *Proc. Natl. Acad. Sci. U.S.A.* 73:375-380.
- Palmiter, R. D. 1977. Prevention of NH₂-terminal acetylation of proteins synthesized in cell-free systems. *J. Biol. Chem.* 252:8781-8783.
- Penman, S., K. Scherrer, Y. Becker, and J. E. Darnell. 1963. Polyribosomes in normal and poliovirus-infected HeLa cells and their relationship to messenger RNA. *Proc. Natl. Acad. Sci. U.S.A.* 49:654-661.
- 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.
- 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, p. 131-213. *In* H. Fraenkel-Conrat and R. R. Wagner (ed.), *Comprehensive virology*, vol. 6. Plenum Press, New York.
- Saborio, J. L., S. S. Pong, and G. Koch. 1974. Selective and reversible inhibition of initiation of protein synthesis in mammalian cells. *J. Mol. Biol.* 85:195-211.
- Sangar, D. V., D. N. Black, D. J. Rowlands, T. J. R. Harris, and F. Brown. 1980. Location of the initiation site for protein synthesis on foot-and-mouth disease virus RNA by *in vitro* translation of defined fragments of RNA. *J. Virol.* 33:59-68.
- Semler, B. L., C. W. Anderson, N. Kitamura, P. G. Rothberg, W. L. Wishart, and E. Wimmer. 1981. Polio-

- rus replication proteins: RNA sequence encoding P3-1b and the sites of proteolytic processing. *Proc. Natl. Acad. Sci. U.S.A.* **78**:3464-3468.
28. 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.
 29. Sharp, P. 1981. Speculations on RNA splicing. *Cell* **23**:643-646.
 30. 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.
 31. 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.
 32. Summers, D. F., J. V. Maizel, and J. E. Darnell. 1967. The decrease in size and synthetic activity of poliovirus poly-
somes late in the infectious cycle. *Virology* **31**:427-435.
 33. 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.
 34. van der Werf, S., F. Bregegere, H. Kopecka, N. Kitamura, P. G. Rothberg, P. Kourilsky, E. Wimmer, and M. Girard. 1981. Molecular cloning of the genome of poliovirus type 1. *Proc. Natl. Acad. Sci. U.S.A.* **78**:5983-5987.
 35. Wold, F. 1981. *In vivo* chemical modification of proteins. *Annu. Rev. Biochem.* **50**:783-814.
 36. Wouters, M., and J. Vandekerckhove. 1976. Amino acid composition of the poliovirus capsid polypeptides isolated as fluorescamine conjugates. *J. Gen. Virol.* **33**:529-533.
 37. 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.