JOURNAL OF VIROLOGY, Feb. 1982, p. 726–729
0022-538X/82/020726-04$02.00/0

Vol. 41, No. 2

Identification of the Polyprotein Termination Site on Encephalomyocarditis Viral RNA

NEIL L. DRAKE,* ANN C. PALMENBERG, AMIT GHOSH, DANIEL R. OMILIANOWSKI, AND PAUL KAESBERG

Biophysics Laboratory of the Graduate School and Biochemistry Department of the College of Agriculture and Life Sciences, University of Wisconsin, Madison, Wisconsin 53706

Received 18 June 1981/Accepted 9 September 1981

We show by sequence analysis of a 420-base-long region adjacent to the 3′ polyadenylic acid of encephalomyocarditis viral RNA and by carboxy terminus analysis of protein E that the termination site of encephalomyocarditis virus polyprotein translation consists of two successive UAG codons located at positions 121 to 126 from the 3′ polyadenylic acid.

The genome of encephalomyocarditis virus (EMC), like that of other picornaviruses, is a single strand of infectious, messenger-active RNA nearly 7,600 bases long (12). Almost all of the genome is devoted to the coding of a polyprotein precursor molecule (molecular weight, 2.4×10^5) from which mature viral proteins are derived in a series of proteolytic cleavage reactions (10).

Sequence analysis of the 3′-terminal region and amino acid analysis of EMC protein E have allowed us to identify the termination site of the polyprotein on the EMC genome. We show below that the termination site is two successive UAG codons at 121 to 126 bases from the 3′-terminal polyadenylic acid [poly(A)], and we list the next 300 bases corresponding to the last 98 amino acids of the polyprotein.

The RNA was sequenced by the dideoxy method of Sanger et al. (11) adapted for use with RNA (1, 15). The 420-base sequence adjacent to the 3′ poly(A) is shown in Fig. 1. This sequence includes a previously reported shorter sequence for this region (15) and resolves several ambiguities that arose in the previous study due to band compression.

Translation termination codons within our sequence are also shown in Fig. 1. Only the 5′-proximal termination codons for each of the three reading frames are possible as termination sites for the polyprotein. These codons are UGA at positions 395 to 397 (frame 1), UGA at positions 366 to 368 (frame 2), and the double terminators UAGUAG at positions 121 to 126 (frame 3).

The partial protein sequences for the carboxy termini of the polyproteins predicted by these termination codons are shown in Fig. 2. Of the three possibilities, phase 3 contains the longest open reading frame within our sequence. Moreover, it is also most closely analogous in position to the double terminators reported for poliovirus (67 to 72 bases from the 3′ poly(A) (7). We therefore thought it likely that our double codons at positions 121 to 126 represented the termination site of EMC translation.

To test this hypothesis, we used selective digestion with carboxypeptidases of defined cleavage specificities to sequentially release amino acids from the carboxy terminus of the labeled EMC protein E. We prepared protein E (9), which was labeled in vitro with [^{14}C]arginine and [^{3}H]tryptophan. Protein E, a viral replicase-related protein (14), has been shown by pactamycin experiments and tryptic peptide mapping to be derived from the carboxy terminus of the polyprotein (4, 5, 9, 14). If phase 3 is the authentic translation phase, then protein E would be expected to have tryptophan and arginine residues (Fig. 2, asterisks) near its carboxy terminus. Specifically, the C-terminal residue and the no. 6 amino acid residue from the C terminus should be tryptophan. Arginines should occupy positions 5 and 7. The predicted proteins for phases 1 and 2 do not have arginine and tryptophans similarly disposed.

The results of these experiments are shown in Fig. 3. The addition of carboxypeptidase A to protein E resulted in the immediate release of [^{3}H]tryptophan. The reaction was 40% complete within 10 s and 100% complete within 2 min. The rapid rate of this reaction indicated that tryptophan is located at, or very near, the carboxy terminus of the protein.

Since the specificity of carboxypeptidase A does not include arginine (2, 3), the released tryptophan must be located on the carboxy side of any arginine-containing sequence. Only the predicted protein for phase 3 contains a tryptophan in such a position. The phase 3 sequence predicts that progressive cleavage by carboxypeptidase A should be blocked by the Arg-Trp-

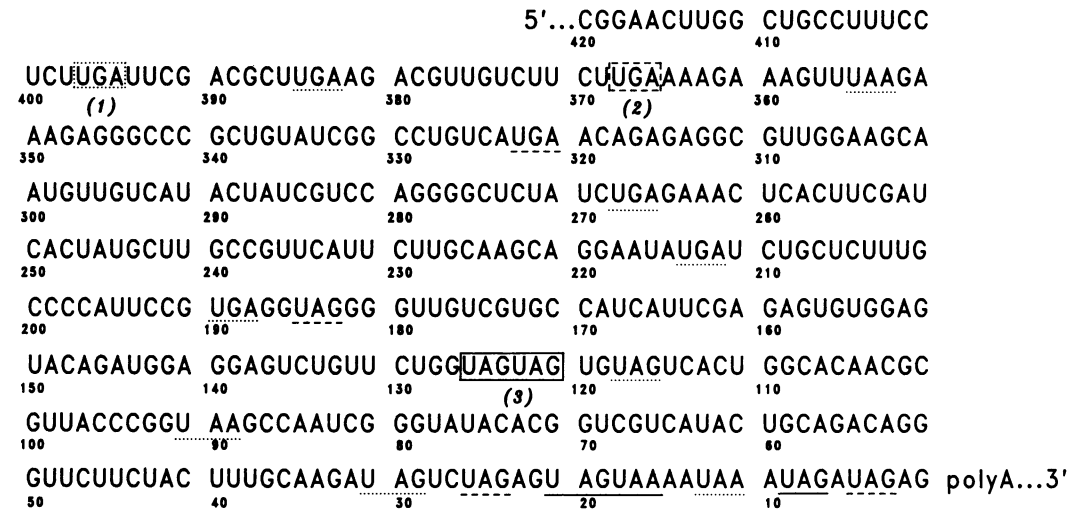

FIG. 1. The 3'-terminal nucleotide sequence of EMC. Sequencing procedures were as previously described (1, 15). Primer (pdT_8dC, PL Biochemicals) was labeled by using [γ-^{32}P]ATP and purified by two-dimensional gel electrophoresis (6, 13). Sequencing reaction mixtures were fractionated on 40-cm-long gels containing 8 or 12% polyacrylamide or, alternatively, on 85-cm-long gels containing 6% polyacrylamide. All gels were 0.35 mm thick. Buffer was used at one-half strength with 85-cm-long gels. Autoradiography was performed with and without intensifying screens (8). Sequencing procedures were repeated several times to generate overlapping band patterns which could be unambiguously read from the gels. All termination codons are highlighted (phase 1 [····], phase 2 [---], phase 3 [———]). The 5'-most termination codon for each phase is boxed.

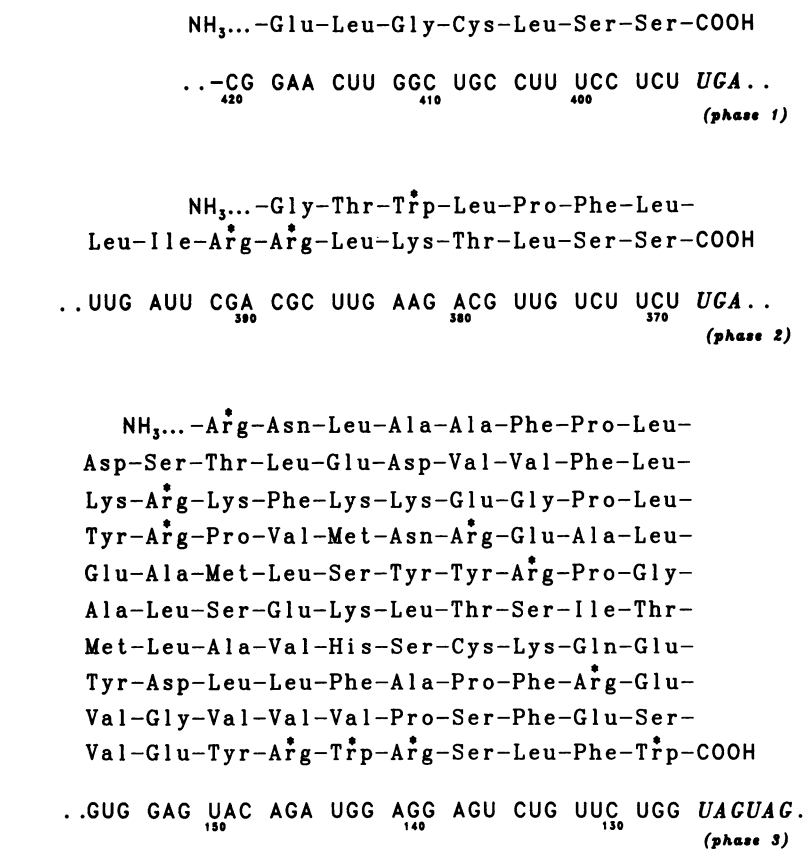

FIG. 2. Potential amino acid sequences for the carboxy terminus of EMC protein E, derived from the three reading frames. Arginine and tryptophan residues are indicated with asterisks (*).

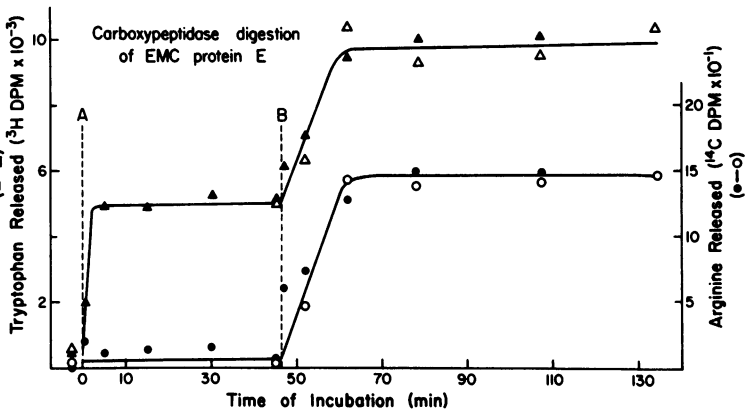

FIG. 3. Carboxypeptidase digestion of protein E. EMC RNA (2 μg) was translated in an extract from rabbit reticulocytes (60 μl) containing [^{3}H]tryptophan (500 μCi at 26 Ci/mmol) and [^{14}C]arginine (100 μCi at 334 mCi/mmol). Viral protein E was isolated by preparative electrophoresis as previously described (9). The final preparation contained (200 μl): 1.7 × 10^6 dpm of ^{3}H, 8.6 × 10^3 dpm of ^{14}C, and 50 μg of bovine serum albumin carrier in 0.2 M morpholine acetate buffer, pH 8.5. A sample of protein E (100 μl) was treated with sodium dodecyl sulfate (5 μl, 10%) and then brought to 37°C in a water bath. An aliquot (10 μl) was removed as a predigestion control and mixed with cold 10% trichloroacetic acid (125 μl). Carboxypeptidase A (40 μl, 1 mg/ml, 50 U/mg, Sigma Chemical Co.) prepared as described before (2) was added to the remaining sample, and at the indicated times, aliquots (10 μl) were removed and treated with trichloroacetic acid as described above. After 45 min at 37°C, carboxypeptidase B (20 μl, 0.5 mg/ml, 100 U/mg, Sigma) was added to the remaining reaction mixture. Further aliquots (10 μl) were withdrawn at the indicated times and treated with trichloroacetic acid. When all samples were collected, bovine serum albumin (50 μg) was added to each sample, and then the insoluble material was pelleted by centrifugation at 15,000 × *g* for 5 min. The clarified supernatant (100 μl) was counted by liquid scintillation for released ^{3}H and ^{14}C. All values were corrected for counting efficiency, isotope spillover, and dilution effects by enzyme addition. The data are expressed as total disintegrations released per minute during the reaction. The open (○, △) and closed (●, ▲) symbols represent the results of two separate experiments. The vertical dashed lines indicate the times of addition of carboxypeptidases A and B.

Arg sequence 5 to 7 residues from the end. If this sequence is correct, then removal of the terminal arginine should allow subsequent release of a second tryptophan. Thus, carboxypeptidase B, which will release arginine (2, 3), was added to the reaction mixture after the digestion with carboxypeptidase A. The addition of this enzyme after the carboxypeptidase A reaction caused rapid release of both tryptophan and arginine (Fig. 3). If we assume that the ^{3}H liberated by the A enzyme represented one tryptophan molecule per molecule of protein E, then reaction with the B enzyme also released one tryptophan residue. It is unclear whether the released ^{14}C represented one or two arginines. Although our data cannot rule out the possibility that the polyprotein terminates at somewhere 5′ to our sequence, the probability of a protein having the amino acids arginine and tryptophan very close to its carboxy terminus and in the order demanded by our experimental results is small. We thus conclude that EMC polyprotein translation terminates at the double codons UAGUAG, 121 to 126 bases from the 3′ poly(A) of the viral RNA.

Comparisons of our EMC RNA sequence and deduced phase 3 amino acid sequence for protein E were made with those of the corresponding regions of polio type I RNA and the amino acid sequence of the homologous protein NCVP4. RNA sequence homology was sought by making computer-generated matrix plots of the two sequences. The analysis revealed no striking homology. This may be seen in Fig. 4, which shows all regions in which at least 20 of 36 bases correspond. There are nine such regions, apparently randomly distributed, with no evidence that they form a pattern of homology (e.g., several such regions forming a long sequence). The longest region of perfect homology found was a 10-base sequence (region A of Fig. 4) from positions 58 to 67 in EMC that is identical to a polio sequence at 32 to 41 bases from the poly(A). An 8-base identical match occurred between EMC bases 365 to 372 and polio bases 331 to 338 (region B of Fig. 4).

Comparison of the deduced amino acid sequence for EMC protein E with that of the corresponding region of polio protein NCVP4 showed one region of sequence similarity. Within EMC, amino acids 77, 79, 80, 81, 82, 84, 86, and 88 (from the carboxy terminus) matched

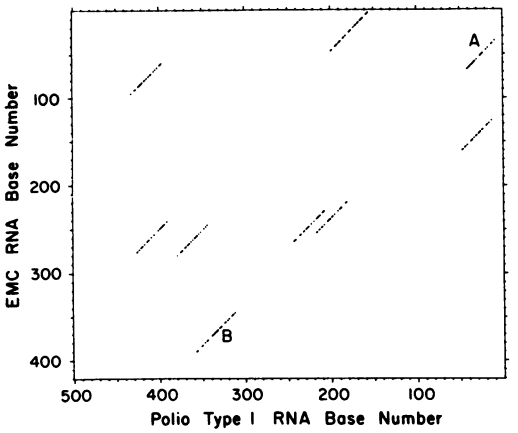

FIG. 4. Computer-generated matrix plot showing sequence homology between polio type 1 RNA and EMC RNA. The polio type 1 RNA sequence on the horizontal axis [numbers indicate the base number from the 3' poly(A)] was compared with the EMC RNA sequence on the vertical axis. The computer compared sequential 36 base segments of the two sequences and searched for a minimum of 20 matches in the segments. When the search was positive, the positions of the matching bases were indicated by dots. Matching segments marked A and B are described in the text. Perfect homology would be indicated by a continuous line of dots starting in the upper right hand corner and extending downward to the left at a 45° angle.

residues 84, 86, 87, 88, 89, 91, 93, and 95 of protein NCVP4, i.e., a match of 8 out of 12 amino acids. These are the amino acids coded by region B of Fig. 4.

The limited RNA and protein sequence homology found is surprising in view of the extensive physical and biological similarities between polio and EMC (12). Whether greater homology exists elsewhere in the two genomes awaits determination of their complete sequences.

Avian myeloblastosis virus reverse transcriptase was the generous gift of J. W. Beard and the Office of Program Resources and Logistics, National Cancer Institute. EMC RNA was kindly provided by Mark Pallansch.

This research was supported by the National Institutes of Health under Public Health Service grants AI-1466, AI-15342, and AI-17331, Career Award AI-21942, and training grant GM07215.

LITERATURE CITED

1. **Ahlquist, P., R. Dasgupta, and P. Kaesberg.** 1981. Near identity of 3' RNA secondary structure in bromoviruses and cucumber mosaic virus. Cell **23**:183–189.
2. **Ambler, R. P.** 1972. Enzymatic hydrolysis with carboxypeptidases. Methods Enzymol. **25**:143–154.
3. **Ambler, R. P.** 1972. Carboxypeptidases A and B. Methods Enzymol. **25**:262–272.
4. **Butterworth, B. E., L. Hall, C. M. Stoltzfus, and R. R. Rueckert.** 1971. Virus-specific proteins synthesized in encephalomyocarditis virus-infected HeLa cells. Proc. Natl. Acad. Sci. U.S.A. **68**:3083–3087.
5. **Butterworth, B. E., and R. R. Rueckert.** 1972. Gene order of encephalomyocarditis virus as determined by studies with pactamycin. J. Virol. **9**:823–828.
6. **DeWachter, R., and W. Fiers.** 1972. Preparative two-dimensional polyacrylamide gel electrophoresis of ^{32}P-labeled DNA. Anal. Biochem. **49**:184–197.
7. **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.
8. **Laskey, R. A.** 1980. The use of intensifying screens or organic scintillators for visualizing radioactive molecules resolved by gel electrophoresis. Methods Enzymol. **65**:363–371.
9. **Palmenberg, A. C., M. Pallansch, and R. R. Rueckert.** 1979. Protease required for processing picornaviral coat protein resides in the viral replicase gene. J. Virol. **32**:770–778.
10. **Rueckert, R. R., T. J. Matthews, O. M. Kew, M. Pallansch, C. McLean, and D. Omilianowski.** 1979. Synthesis and processing of picornaviral polyprotein, p. 113–126. *In* R. Perez-Bercoff (ed.), Molecular biology of picornaviruses. Plenum Publishing Corp., New York.
11. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
12. **Scraba, D. G.** 1979. The picornavirion: structure and assembly, p. 123. *In* R. Perez-Bercoff (ed.), Molecular biology of picornaviruses. Plenum Publishing Corp., New York.
13. **Symons, R. H.** 1979. Extensive sequence homology at the 3'-termini of the four RNAs of cucumber mosaic virus. Nucleic Acids Res. **7**:825–837.
14. **Traub, A., B. Diskin, H. Rosenberg, and E. Kalmar.** 1976. Isolation and properties of the replicase of encephalomyocarditis virus. J. Virol. **18**:375–382.
15. **Zimmern, D., and P. Kaesberg.** 1978. 3'-terminal nucleotide sequence of encephalomyocarditis virus RNA determined by reverse transcriptase and chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **75**:4257–4261.