Partial N-terminal amino acid sequences of polypeptides p14 and p12 of encephalomyocarditis virus are identical and correspond to the N-terminus of the viral polyprotein

T.Yu. Ugarova, E.Yu. Siyanova, Yu.V. Svitkin*, Yu.A. Kazachkov*, L.A. Baratova and V.I. Agol*+

A.N. Belozersky Laboratory of Molecular Biology and Bioorganic Chemistry, Moscow State University, W-234, Moscow 119899 and **Institute of Poliomyelitis and Viral Encephalitides, the USSR Academy of Medical Sciences, Moscow Region 142782, USSR*

Received 14 March 1984; revised version received 30 March 1984

Our previous data suggested that translation in an EMC virus RNA-programmed cell-free system from Krebs-2 cells is initiated predominantly at a single site and that the earliest amino acid sequences synthesized correspond to non-structural 'leader' polypeptides p14 and p12 [(1982) FEBS Lett. 141, 153–156]. Here, polypeptides p14 and p12 were labelled in vitro by tritiated amino acids, isolated and subjected to automated Edman degradation. Both polypeptides (after the loss of the N-terminal methionine) were shown to contain alanine in position 1 and glutamic acid in positions 5 and 7. These and other data demonstrate that p14 and p12 share a common N-terminal sequence. This sequence coincides precisely with the N-terminus of EMC virus polyprotein sequence deduced from the primary structure of the viral genome [(1984) Nucleic Acids Res., in press]. Thus, the single initiation site operating in our translation system corresponds to the start of the polyprotein molecule.

Picornavirus *Cell-free protein synthesis* *Translation initiation* *Polyprotein processing* *Radiochemical sequence analysis*

1. INTRODUCTION

All the proteins of picornaviruses found in the infected cell are known to be generated by multistep processing of a giant polyprotein molecule, which is initiated at a single site on the viral genome (review [1]). For poliovirus, this site corresponds precisely to the 5′-terminus of the genes coding for capsid proteins [2–4]. On the other hand, in vitro translation of the genome of another picornavirus, encephalomyocarditis (EMC) virus, appears to be initiated at a site(s) other than the beginning of the genes for capsid proteins [5–8]. Upon translation of EMC virus RNA in extracts from Krebs-2 cells in the presence of formyl-[^{35}S]methionyl-tRNA$_f^{Met}$, the label was detected first in a high-M_r polypeptide, preA, and then, upon subsequent processing of preA, in two small polypeptides with apparent M_r values of 14000 (p14) and 12000 (p12) [7]. The amino acid sequences of p14 and p12, being closely related, appear to precede the sequences of capsid proteins in the preA molecule, therefore, p14 and p12 were designated as leader polypeptides.

The existence of two related N-terminal polypeptides may be explained in two different ways:

(i) p14 and p12 are initiated at the same site but either both are cleaved from preA at different points, or one of them undergoes a modification that changes its electrophoretic mobility;
(ii) p14 and p12 are initiated at distinct in-frame initiation codons. The first possibility implies that p14 and p12 share a common N-terminal

+ To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

amino acid sequence, whereas the second one predicts that the N-termini of the two polypeptides are dissimilar.

Our preliminary analyses failed to reveal any differences between the N-terminal tryptic peptides of p14 and p12 in terms of their electrophoretic and chromatographic behaviour [7,9]. However, we felt it desirable to compare the N-termini of p14 and p12 by a more direct method. Partial N-terminal amino acid sequences of these polypeptides have therefore been determined and found to be identical. The results strongly suggest that only a single initiation codon is used for the synthesis of EMC virus polyprotein. Since the primary structure of the EMC virus RNA has recently been determined [10], we are able to locate unambiguously this initiation codon on the viral genome.

2. MATERIALS AND METHODS

RNA isolated from purified preparations of EMC virus [11] was translated in micrococcal nuclease-treated extracts from Krebs-2 cells [12] at 130 mM KCl and 3.3 mM $MgCl_2$. For the N-terminal labelling of newly synthesized polypeptides, formyl-[^{35}S]methionyl-tRNA$_f^{Met}$ [13] was used as a substrate. The labelling of internal amino acid residues was achieved by using one of the two mixtures of tritiated amino acids (produced by the Institute of Molecular Genetics, Moscow). Mixture I contained [^{3}H]alanine (64 Ci/mmol), [^{3}H]aspartic acid (32 Ci/mmol), [^{3}H]leucine (80 Ci/mmol) and [^{3}H]valine (58 Ci/mmol); it was used at a final concentration of 20 μM. Mixture II contained [^{3}H]glutamic acid (17 Ci/mmol), [^{3}H]isoleucine (63 Ci/mmol), [^{3}H]phenylalanine (19 Ci/mmol) and [^{3}H]proline (80 Ci/mmol); it was used at 28 μM. Both mixtures also contained [^{35}S]methionine (~9 Ci/mmol); the final concentration of this amino acid in the incubation mixes was 10 μM. The concentration of unlabelled amino acids was 40 μM. The translation samples (0.4 ml) were incubated for 30 min at 30°C, further elongation was stopped by the addition of cycloheximide to a final concentration of 100 μg/ml, and 20 μl of a high-speed supernatant (fraction S_{150}) from EMC virus-infected Krebs-2 cells were added as a source of the virus-specific protease [7,14]. After an additional incubation for 5 min at 30°C, the samples were fixed by the addition of 2 volumes of the dissociating buffer [12] followed by heating at 100°C for 5 min.

The translation products were separated by electrophoresis on SDS-containing polyacrylamide gel slabs [15]. The labelled bands were excised, washed with methanol, swollen in water and homogenized in 70% formic acid at 0°C. The elution continued for 5 h, the particulate material was pelleted and re-extracted twice with 60% formic acid for 2 h. The eluted proteins exhibited an unexpectedly low level of ^{35}S-radioactivity (due perhaps to a relatively long storage of [^{35}S]methionine before use). This fact precluded reliable location of the internal methionine residues in the sequences reported below.

Labelled proteins in the eluates were mixed with apomyoglobin and immediately subjected to automated Edman degradation in a Beckman Sequencer, Model 890C. Aliquots of each sequence cycle sample were supplemented with 3-phenyl-2-thiohydantoin (PTH) amino acid standards and subjected to high performance reverse-phase liquid chromatography on Ultrasphere ODS-C18, 5 μm (Altex). The labelled PTH amino acid derivatives were identified by their co-elution with the standards.

3. RESULTS AND DISCUSSION

Here, p14 and p12 were generated by treatment of a mixture of high-M_r precursor polypeptides with a preparation of virus-specific protease [7,14] (fig.1).

When mixture I was used as a donor of radioactive amino acids, [^{3}H]alanine was identified at position 1 of both p14 and p12; with mixture II, [^{3}H]glutamic acid was found at positions 5 and 7 of both polypeptides (fig.2). Furthermore, as evidenced by formyl-[^{35}S]methionine-labelling (fig.1a), the amino acid sequence of p14 and p12 should be started with a methionine residue (the N-terminal methionine, in contrast to formylmethionine, is known to be often cleaved from newly synthesized polypeptides; this explains why the determined sequence of the leaders does not begin with a methionine). Thus, 4 positions out of 9 investigated are demonstrated to be occupied by the same amino acid residues in p14 and p12, respectively. Moreover, none of the remaining 6

Fig.1. Preparative electrophoresis of polypeptides p14 and p12. Cell-free translation samples were incubated for 30 min in the presence of EMC virus RNA and either f-[^{35}S]met-tRNA$_f^{Met}$ (a), or radioactive amino acid mixture I (b); then cycloheximide and a crude preparation of the EMC virus-specific protease were added, and incubation was continued for 5 min. For other details, see section 2. Only parts of the preparative slab gels are shown.

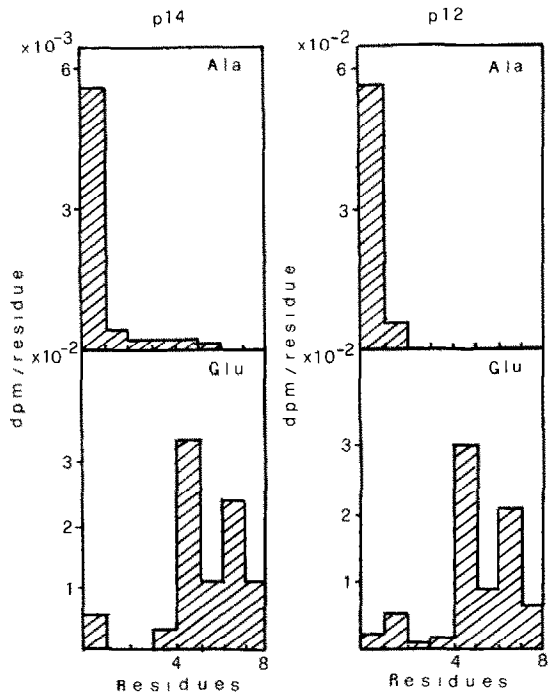

Fig.2. Determination of the partial N-terminal amino acid sequences of polypeptides p14 and p12. These polypeptides were synthesized in vitro in the presence of either mixture I (top panels) or mixture II (bottom panels) and separated by electrophoresis as in fig.1. The radioactive samples after each cycle of Edman degradation were chromatographed as described in section 2. Within the N-terminal octapeptide, labelled alanine and glutamic acid were only detected in the polypeptides synthesized in the presence of mixture I and mixture II, respectively.

^{3}H-labelled amino acids that were contained in mixtures I and II, appear to be present at positions 1–9 of the p14 or p12 amino acid sequences (not shown). Taking into account the close similarity of the two leader polypeptides in terms of their electrophoretic, chromatographic and gel-filtration behaviour [7,9], we can safely conclude that p14 and p12 share a common N-terminal amino acid sequence and hence are started at the same initiation codon.

Now that the primary structure of the entire coding segment of the EMC genome has been determined [10], this initiation codon can be identified. It corresponds to the first AUG triplet of the large open reading frame coding for the polyprotein (fig.3). Indeed, the N-termini of the polypro-

```
1. AUG GCC ACA ACC AUG GAA CAA GAG ACU
2. Met Ala Thr Thr Met Glu Gln Glu Thr
3. Met Ala -   -   -   Glu -   Glu -
```

Fig.3. Nucleotide sequences of the 5′-end of the large open reading frame of the EMC virus genome (upper line) [10], deduced N-terminal amino acid sequence of the EMC virus polyprotein (middle line) [10] and partial N-terminal amino acid sequence of p14 and p12 synthesized in vitro (lower line).

tein as well as of p14 and p12 share the following features: the 3 polypeptides are started with a methionine residue; after the loss of the N-terminal methionine, the 3 polypeptides contain an alanine residue at the 1st position and glutamic acid residues at the 5th and 7th positions; and neither contains aspartic acid, isoleucine, leucine, phenylalanine, proline or valine at positions 1–8. No other region exists in the polyprotein molecule [10] with this combination of properties. Moreover, a small open reading frame located upstream from the polyprotein-coding sequence (A. Palmenberg, personal communication, 1983), if translated at all, will direct the synthesis of a polypeptide with a different N-terminal structure. The identity of the N-terminal amino acid sequences of the p14 (p12) and of the polyprotein is also supported by the following fact: the M_r of the N-terminal tryptic peptide of the leaders was estimated by gel filtration to be in the range of 2500–3500 [9], in reasonable accord with the value of 2417 calculated from the sequencing data [10].

The primary structure of the EMC virus genome also confirms the existence of a leader amino acid sequence, which precedes the capsid polypeptides in the polyprotein; the predicted M_r of this leader is about 8000 [10], a value somewhat lower than our previous estimates for p14 and p12 [7]. However, since these two polypeptides exhibit anomalous electrophoretic behaviour in SDS-containing polyacrylamide gels (depending on conditions, the leaders migrate either slower or faster than EMC virus-specific polypeptides H and I; unpublished), we now consider it very likely that the actual M_r of p14 or p12 is about 8000. Knowledge of the C-terminal structure of p14 and p12 is needed to determine their size precisely and to elucidate the difference between the 2 polypeptides.

Summing up, we can state that translation of the EMC virus genome in our cell-free system is initiated predominantly at a single site corresponding to the start of the polyprotein molecule.

ACKNOWLEDGEMENT

We wish to thank Dr Ann C. Palmenberg who made available the primary structure of the EMC virus genome prior to publication.

REFERENCES

[1] Rueckert, R.R., Matthews, T.J., Kew, O.M., Pallansch, M., McLean, C. and Omilianowski, D. (1979) in: The Molecular Biology of Picornaviruses (Perez-Bercoff, R. ed) pp.113–125.
[2] Kitamura, N., Semler, B.L., Rothberg, P.G., Larsen, G.R., Adler, C.J., Dorner, A.J., Emini, E.A., Hanecak, R., Lee, J.J., Van der Werf, S., Anderson, C.W. and Wimmer, E. (1981) Nature 291, 547–553.
[3] Racaniello, V.R. and Baltimore, D. (1981) Proc. Natl. Acad. Sci. USA 78, 4887–4891.
[4] Dorner, A.J., Dorner, L.F., Larsen, G.R., Wimmer, E. and Anderson, C.W. (1982) J. Virol. 42, 1017–1028.
[5] Smith, A.E. (1973) Eur. J. Biochem. 33, 301–313.
[6] Lawrence, C. and Thach, R.E. (1975) J. Virol. 15, 918–928.
[7] Kazachkov, Yu.A., Chernovskaya, T.V., Siyanova, E.Yu., Svitkin, Yu.V., Ugarova, T.Yu. and Agol, V.I. (1982) FEBS Lett. 141, 153–156.
[8] Campbell, E.A. and Jackson, R.A. (1983) J. Virol. 45, 439–441.
[9] Siyanova, E.Yu., Chernovskaya, T.V., Svitkin, Yu.V., Kazachkov, Yu.A. and Ugarova, T.Yu. (1984) Molek. Genet. Mikrobiol. i Virusol. (Moscow) no.2, 41–45.
[10] Palmenberg, A.C., Kirby, E.M., Janda, M.R., Drake, N.L., Duke, G.M., Potratz, K.F. and Collett, M.S. (1984) Nucleic Acids Res., in press.
[11] Kerr, I.M. and Martin, E.M. (1972) J. Virol. 9, 559–561.
[12] Svitkin, Yu.V. and Agol, V.I. (1978) FEBS Lett. 87, 7–11.
[13] Svitkin, Yu.V., Ugarova, T.Yu., Chernovskaya, T.V., Lyapustin, V.N., Lashkevich, V.A. and Agol, V.I. (1981) Virology 110, 26–34.
[14] Gorbalenya, A.E., Svitkin, Yu.V., Kazachkov, Yu.A. and Agol, V.I. (1979) FEBS Lett. 108, 1–5.
[15] Laemmli, U.K. (1970) Nature 227, 680–685.
[16] Gorbalenya, A.E., Svitkin, Yu.V. and Agol, V.I. (1981) Biochem. Biophys. Res. Commun. 98, 952–960.
[17] Gorbalenya, A.E. and Svitkin, Yu.V. (1983) Biokhimiya 48, 442–453.