POLYPROTEIN PROCESSING OF ALPHAVIRUSES: N-TERMINAL STRUCTURAL ANALYSIS OF SEMLIKI FOREST VIRUS PROTEINS p62, E3 AND ns70

Nisse KALKKINEN, Hans JÖRNVALL[+] and Leevi KÄÄRIÄINEN*

*Department of Biochemistry, University of Helsinki, 00170 Helsinki 17, *Department of Virology, University of Helsinki, 00290 Helsinki 29, Finland* and *[+]Department of Chemistry I, Karolinska Institutet, 104 10 Stockholm 60, Sweden*

Received 19 January 1981

1. Introduction

The genome of Semliki Forest virus (SFV) apparently codes for 4 structural and 4 non-structural proteins.

The structural proteins are translated from a single initiation site essentially as a polyprotein. The gene order in the subgenomic 26 S RNA messenger is as follows: capsid protein–E3–E2–E1. The individual proteins are formed by post-translational processing of the structural polyprotein in different compartments of the cell (reviewed [1,2]), [3–5].

The non-structural proteins are also synthesized from one initiation site, principally as a polyprotein [5] from the 42 S RNA genome which serves as messenger. The gene order in the non-structural polyprotein is known to be ns70–ns86–ns72–ns60 (reviewed [1,2]). The details of the post-translational processing of the non-structural polyprotein are unknown.

The nucleotide sequence of the cloned SFV 26 S RNA has been determined [3,4]. From this, the protein structures could be deduced using the partially known N- and C-terminal protein amino acid sequences determined in [5–8] and here.

We report the results of direct N-terminal sequence analysis of the structural precursor protein p62 (E3 + E2) and the envelope glycoprotein E3 which allow their precise localization in the polyprotein. Both proteins have identical N-terminal structures which means that the signal sequence of p62 is not cleaved off as is the case with most other signal sequences. A mixture of radiolabelled p62 and ns70 gave positive assignments only in positions corresponding to those of p62 suggesting that ns70 may have a blocked N-terminus.

2. Materials and methods

Semliki Forest virus prototype strain was grown, labelled with [^{35}S]Met and purified as in [6]. After purification of E3 glycoprotein as in [6], it (20 nmol) was treated with 0.4 μg neuraminidase (Behringewerke AG, Marburg) in 0.1 M sodium acetate (pH 5.5), 1 mM $CaCl_2$, at 37°C for 2 h, and then chromatographed on a Bio-Gel P2 column (0.7 × 15 cm, 10% acetic acid) and lyophilized. For quantitation, a sample (5%) of the purified E3 was subjected to hydrolysis in 6 N HCl at 110°C for 24 h followed by amino acid determination (Beckman 121M amino acid analyzer).

Virus-specific proteins were labelled in BHK 21 (baby hamster kidney) cells infected with the wild-type virus or in secondary, pathogen-free chicken embryo cells infected with the *ts*-1 mutant as in [9]. At 5.5 h post-infection the cells were labelled with [^{3}H]Ala (43 Ci/mmol, Amersham), 500 μCi/dish in Eagle's minimum essential medium (MEM). At 6 h post-infection the dishes were once washed with Hank's balanced salt solution and the cells were collected into 2% sodium dodecyl sulfate (SDS) as in [9].

Infected chicken embryo cells were, after synchronization of initiation of protein synthesis [9,10], labelled at 5.5 h post-infection for 3 min with either [^{3}H]Ala (43 Ci/mmol, Amersham), [^{3}H]Val (57 Ci/mmol, Amersham) or [^{35}S]Met (210 Ci/mmol, Amersham), 500 μCi/dish in MEM in which the amino acid used for labelling was lacking. The pulse was followed by a 1 h chase in the presence of excess of the respective amino acid in MEM. The cells were collected into 2% SDS as in [9].

Preparative SDS–polyacrylamide gel electrophoresis was carried out in the presence of 2-mercaptoethanol in a 10% gel and the proteins were localized by the aid of radioactive markers run in separate lanes followed by elution from the gel and precipitation as in [6].

The precipitated proteins were dissolved in 0.5 ml formic acid and degraded in a Beckman model 890C sequencer together with 78 nmol apo-myoglobin and 4 mg Polybrene [6]. The 1 M Quadrol protein program was used. After conversion (0.2 ml 1M HCl, 80°C, 5 min) and extraction (2 × 0.7 ml ethyl acetate), the dried ethyl acetate phases were dissolved in 60 μl methanol and 50 μl from each fraction was directly measured for radioactivity. The remaining 10 μl was analyzed by high-performance liquid chromatography (HPLC, Hewlett-Packard 1084B) using a Nucleosil ODS C_{18} column and a solvent system as in [11].

Purified E3 was manually degraded [12]. After conversion and extraction (as above) the dried ethyl acetate phases were dissolved in 60 μl methanol; 20 μl was directly measured for radioactivity and the remaining 40 μl was analyzed by HPLC as above.

3. Results and discussion

SFV-specific proteins were labelled and then separated by preparative SDS–polyacrylamide gel electrophoresis. Fig.1A shows the separation of [^{3}H]Ala-labelled proteins from wild-type virus-infected BHK 21 cells. During the conditions used, only negligible amounts of virus-specific non-structural proteins were synthesized, as shown by the level of ns86.

After purification the [^{3}H]Ala labelled p62 was subjected to sequencer degradation together with the apo-myoglobin carrier. The results are shown in fig.2A. They allow precise localization of the N-terminus of p62 on the structural polyprotein as shown in table 1. The level of initial coupling for p62 was found to be about the same as that for the apo-myoglobin carrier, indicating that the N-terminus of p62 is accessible for Edman degradations.

However, the finally processed envelope glycoprotein E3 may have a blocked N-terminus as suggested in [6]. Some amino acid derivatives were recovered but only in too low amounts for reliable sequence assignments. To control these results, E3 was labelled with [^{35}S]Met and purified as before and the yield

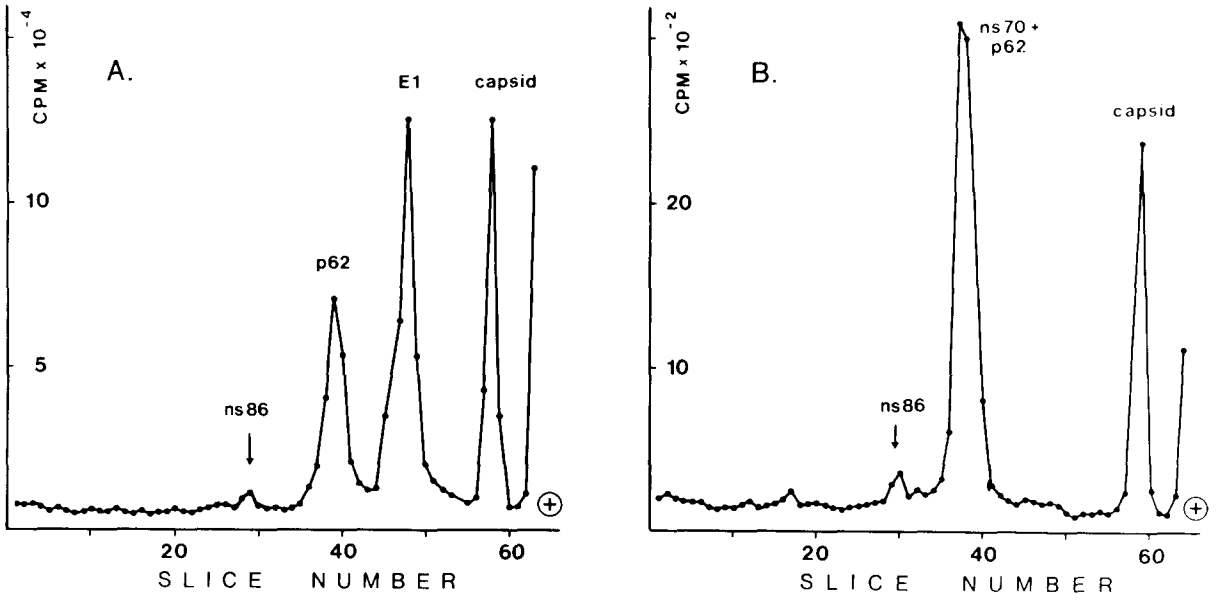


Fig.1. Preparative SDS–polyacrylamide gel electrophoresis of proteins from (A) BHK 21 cells infected with Semliki Forest virus wild-type and labelled at 5.5 h post-infection with [^{3}H]Ala for 30 min. (B) Chicken embryo cells infected with Semliki Forest virus *ts*-1 mutant and labelled, after synchronization of initiation of protein synthesis, at 5.5 h post-infection with a 3 min pulse of [^{3}H]Ala. Slices 63 and 64 indicate positions of bromophenol blue in A and B, respectively.

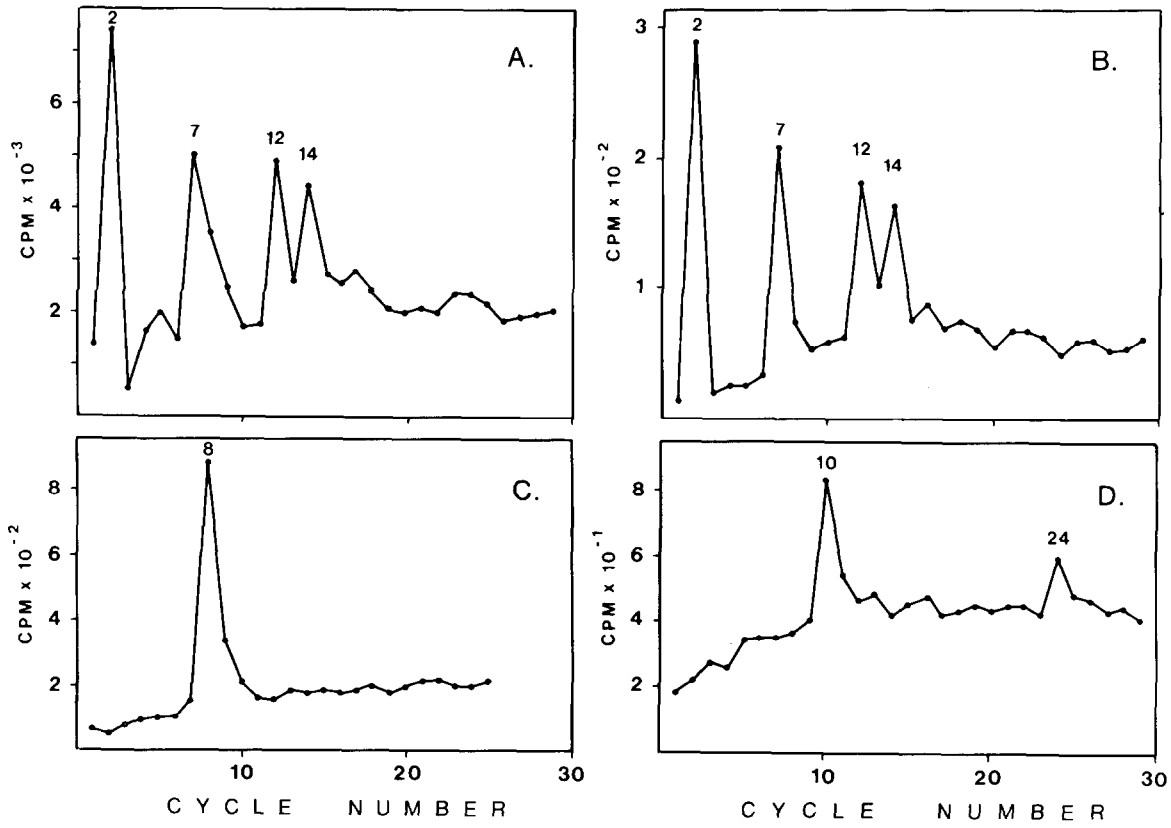
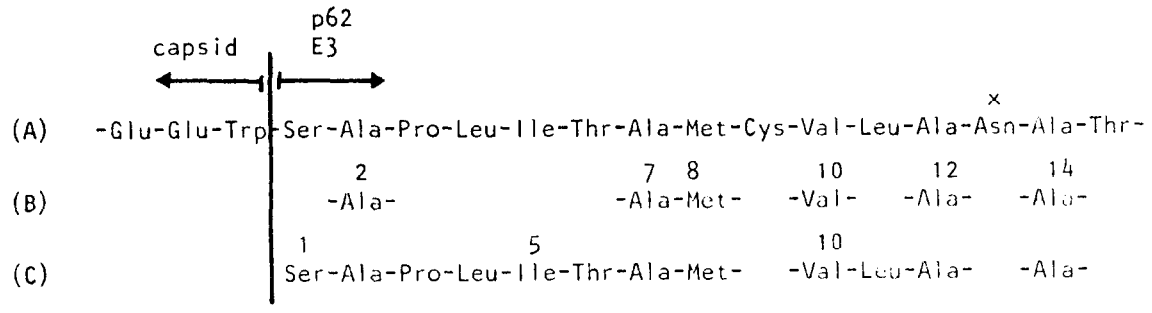


Fig.2. Results of sequencer degradations of purified Semliki Forest virus specific proteins. (A) p62 labelled with [^{3}H]Ala and purified from BHK 21 cells infected with the wild-type virus. (B–D) Mixtures of ns70 and p62 from *ts*-1 mutant infected chicken embryo cells incubated at 39°C and pulse-labelled, after synchronization of initiation of protein synthesis, with [^{3}H]Ala (B), [^{35}S]Met (*C*) and [^{3}H]Val (*D*).

Table 1
Comparison of sequence results with previous data



(A) A part of the indirectly determined [3] amino acid sequence of the Semliki Forest virus (SFV) structural polyprotein around the C-terminus of the capsid protein [7,8]. (B) N-terminal structure presently determined (fig.2) by radiosequence analysis of p62. (C) N-terminal amino acid sequence presently determined (table 2) for the envelope glycoprotein E3. (X) One of the two potential glycosylation sites of SFV E3

Table 2
Results of N-terminal sequence analysis of Semliki Forest virus envelope glycoprotein E3 (10 nmol, 67 000 cpm [^{35}S]Met)

Position	Identified residue	Amount (nmol)	Radioactivity (cpm)	Position	Identified residue	Amount (nmol)	Radioactivity (cpm)
1	Ser H		35	8	Met R		787
2	Ala H	0.21	30	9	–		166
3	Pro H		38	10	Val H	0.23	144
4	Leu H	0.22	36	11	Leu H	0.10	98
5	Ile H	0.24	62	12	Ala H	0.12	74
6	Thr H		47	13	–		92
7	Ala H	0.20	44	14	Ala H	0.10	98

Residue identifications by H (HPLC) or R (radioactivity from the [^{35}S]Met label). Values show levels of the most stable amino acid derivatives recovered

was determined by quantitative amino acid analysis after acid hydrolysis. In contrast to [6] the purified E3 was now treated with neuraminidase to reduce the carbohydrate moiety [13]. E3 (10 nmol) was manually degraded with results shown in table 2. These confirm the structure obtained from the nucleotide sequence [3] and show that both p62 and E3 have an identical N-terminal structure which in the structural polyprotein immediately follows the C-terminus of the capsid protein as shown in table 1. However, again the yield of amino acid derivatives was much lower (table 2) than expected (>5 nmol), although considerably higher than the trace amounts recovered in [6] when sialic acid was not removed. No Asn was recovered in position 13 (table 2), one of the two potential glycosylation sites of E3, suggesting that this residue is linked to carbohydrate. This is in accordance with the preliminary results which show that E3 gives rise to a large carbohydrate-containing tryptic peptide as would be expected if the Asn 13 were glycosylated (Pesonen et al., in preparation).

The quantitative amino acid analysis of E3 (not shown) which was indistinguishable to that calculated from the DNA sequence data [3] was compatible with [7] that the C-terminus of E3 has only one Arg residue. This must mean that upon cleavage of p62 to E3 and E2 one Arg residue of the polyprotein is removed.

The low yields of amino acid derivatives from E3 indicate that its N-terminus is poorly accessible for Edman degradation. The N-terminal structure obtained, Ser–Ala–, is a typical one for acetylation [14]. However, in the case of E3, acetylation is questionable, since substantial amounts apparently remain unblocked. Although incomplete acetylations are known in native systems [15,16], an alternative explanation to the low yields may be chemical hindrance of an unblocked N-terminus by the adjacent glycan at Asn 13. E3 and the bulk of p62 differ in respect to their glycan composition since the latter contains high mannose-type glycans [17] which are smaller. In any event, if the N-terminus of E3 is blocked this must be a late processing since its immediate precursor, p62 is accessible to Edman degradation.

The non-structural polyprotein of SFV has an N-terminal sequence Met–Ala [5]. Since the N-terminal structure of p62 was now known we tried to obtain information on the N-terminal structure of ns70 from a mixture of labelled ns70 and p62 isolated from SFV *ts*-1 mutant-infected chicken embryo cells [18]. SFV-specific non-structural proteins are overproduced in *ts*-1 mutant-infected cells incubated at the restrictive temperature [19]. After synchronization of initiation of protein synthesis [9,10], 3 min pulses of [^{3}H]Ala, [^{3}H]Val or [^{35}S]Met were used to label preferentially ns70 from the N-terminal part of the non-structural polyprotein. Capsid protein and p62 from the structural polyprotein become also labelled in these conditions [9]. Fig.1B shows the preparative separation of the [^{3}H]Ala-labelled proteins. The distribution of radioactivity was essentially the same in the proteins labelled with [^{3}H]Val and [^{35}S]Met (not shown) but the total incorporation was then considerably lower. The negligible amount of labelled ns86 obtained and the lack of the labelled E1 confirmed that synchronization of inititation of protein synthesis had occurred as expected. The [^{3}H]Ala,

[^{3}H]Val and [^{35}S]Met labelled mixtures of ns70 and p62 were subjected to sequencer degradation. Results are shown in fig.2B–D and table 1. All the mixtures contained ≥50% (mol/mol) ns70 as calculated from the radioactivities in relation to the composition of p62. However, the positive assignments in fig.2B–D correspond only to positions in p62 and confirm the results shown in fig.2A and those from the nucleotide sequence data [3] (table 1). No other positive assignments could be obtained, suggesting that the N-terminus of ns70 is also inaccessible to degradation provided that the presently labelled residues (Ala, Val, Met) occur within the 24–29 first positions (fig.2B–D) which they should, especially if the original Met–Ala– [5] or longer oligopeptides have not been removed. Many viral proteins are known to have blocked N-termini and when formed by cleavage of a precursor such proteins are often located at the N-terminus of the original polypeptide [20], ns70 is also N-terminal in the non-structural polyprotein of SFV and thus the present result is not unexpected.

Acknowledgements

We thank Matti Laaksonen for excellent technical assistance and Dr Hans Söderlund for valuable discussions. This investigation was supported by the Finnish National Fund for Research and Development, the Swedish Cancer Society and the Knut and Alice Wallenberg's Foundation.

References

[1] Strauss, J. H. and Strauss, E. G. (1977) in: The molecular biology of animal viruses (Nyak, D. P. ed) pp. 111–166, Dekker, New York.
[2] Kääriäinen, L. and Söderlund, H. (1978) Curr. Top. Microbiol. Immunol. 82, 15–69.
[3] Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H. and Delius, H. (1980) Nature 288, 236–241.
[4] Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H. and Delius, H. (1980) Proc. Natl. Acad. Sci. USA 77, 6376–6380.
[5] Glanville, N., Ranki, M., Morser, J., Kääriäinen, L. and Smith, A. (1976) Proc. Natl. Acad. Sci. USA 73, 3059–3063.
[6] Kalkkinen, J., Jörnvall, H., Söderlund, H. and Kääriäinen, L. (1980) Eur. J. Biochem. 108, 31–37.
[7] Kalkkinen, N. (1980) FEBS Lett. 115, 163–166.
[8] Boege, U., Wengler, G., Wengler, G. and Wittman-Liebold, B. (1980) Virology 103, 178–190.
[9] Lachmi, B. E. and Kääriäinen, L. (1976) Proc. Natl. Acad. Sci. USA 73, 1936–1940.
[10] Saborio, J. L.. Pong, S. S. and Kochi, G. (1974) J. Mol. Biol. 85, 195–221.
[11] Zimmerman, C. L., Appella, E. and Pisano, J. J. (1977) Anal. Biochem. 77, 569–573.
[12] Peterson, J. D., Nehrlich, S., Oyer, P. E. and Steiner, D. F. (1972) J. Biol. Chem. 247, 4866–4871.
[13] Mattila, K., Luukkonen, A. and Renkonen, O. (1976) Biochim. Biophys. Acta 419, 435–444.
[14] Jörnvall, H. (1975) J. Theor. Biol. 55, 1–12.
[15] Terhorst, C., Möller, W., Laursen, R. and Wittman-Liebold, B. (1973) Eur. J. Biochem. 34, 138–152.
[16] Jörnvall, H., Fairwell, T., Kratofil, P. and Wills, C. (1980) FEBS Lett. 111, 214–218.
[17] Personen, M., Saraste, J., Hashimoto, K. and Kääriäinen, L. (1981) Virology, in press.
[18] Keränen, S. and Kääriäinen, L. (1974) Acta Pathol. Microbiol. Scand. Sect B 82, 810–820.
[19] Keränen, S. and Kääriäinen, L. (1975) J. Virol. 16, 388–396.
[20] Bell, J. R. , Hunkapiller, M. W., Hood, L. E. and Strauss, J. H. (1978) Proc. Natl. Acad. Sci. USA 75, 2722–2726.