JOURNAL OF VIROLOGY, June 1981, p. 968-972
0022-538X/81/060968-05$02.00/0

Vol. 38, No. 3

Amino-Terminal Sequence Analysis of Alphavirus Polypeptides

WILLIAM J. WELCH,[1] BARTHOLOMEW M. SEFTON,[1]* AND FRED S. ESCH[2]

Tumor Virology Laboratory[1] and Laboratories for Endocrinology,[2] The Salk Institute, San Diego, California 92138

Received 16 January 1981/Accepted 16 March 1981

The single late 26S mRNA of Semliki Forest virus (SFV) directs the synthesis of the four viral structural proteins, C, E3, E2, and E1, and the recently described nonstructural protein, 6K. We report here partial NH_2-terminal amino acid sequences of the SFV polypeptides E3 and 6K and of p62, the precursor to E3 and E2. In addition, we have determined a partial NH_2-terminal sequence of the Sindbis virus homolog of 6K, the 4.2K protein. p62 and E3 of SFV have identical NH_2-terminal amino acid sequences. Comparison of the partial NH_2-terminal sequences of 6K of SFV and 4.2K of Sindbis virus with the deduced amino acid sequence encoded by the 26S mRNA of each virus reveals that the genes for these peptides are located in each case between those for E2 and E1. The order of the genes on the 26S mRNA of the alphaviruses is therefore 5'-C-E3-E2-6K-E1-3'. We discuss two mechanisms by which the nascent viral glycoproteins may be inserted into the membrane of the endoplasmic reticulum.

The Semliki Forest virus (SFV) virion contains four structural proteins; a nucleocapsid protein, C; and three envelope glycoproteins, E1, E2, and E3 (9). A single subgenomic 26S mRNA directs the synthesis of all four structural polypeptides. The genes for these proteins are arranged on the 26S mRNA in the order of 5'-C-E3-E2-E1-3' (4). Translation is initiated only at a single site near the 5' end of the 26S mRNA, and production of the individual proteins results from a series of proteolytic cleavages of the nascent polypeptide chain (5). Both E3 and E2 are synthesized initially as a precursor polypeptide, p62, which is later cleaved to yield the two mature glycoproteins (15). Newly synthesized capsid protein is released into the cytoplasm and assembles rapidly with the viral genomic 42S RNA to form the viral nucleocapsid. In contrast, the newly synthesized envelope glycoproteins, p62 and E1, are inserted into the membrane of the endoplasmic reticulum and are glycosylated as nascent chains (8, 14). p62 and E1 are subsequently transferred to the plasma membrane of the host cell. Here, the viral nucleocapsid associates with the envelope glycoproteins, and mature virus is released via envelopment of the nucleocapsid by the virally modified portion of the host cell plasma membrane.

We have recently shown that the 26S mRNA of SFV encodes, in addition to the virion structural proteins, a small nonstructural polypeptide of approximately 6,000 daltons which we named the 6K protein (17). Gene mapping studies, using the antibiotic pactamycin, demonstrated that the gene coding for 6K is located downstream from the gene encoding E2, but did not allow determination of whether 6K was to the 5' side or 3' side of the E1 gene (17). Further, it was shown that newly synthesized 6K, like the newly synthesized glycoproteins, appears to be associated with the rough endoplasmic reticulum (RER) in infected cells (17). Both the location of the 6K gene on the 26S mRNA as well as the subcellular location of the 6K protein in the infected cell prompted us to suggest that 6K could function in the insertion of E1 into the membrane of the endoplasmic reticulum. Without a precise location of the 6K gene on the 26S mRNA, however, the role of the 6K protein remained unclear.

The complete nucleotide sequence of the SFV 26S mRNA has recently been determined by Garoff and his colleagues (6, 7). From this sequence it should be possible to deduce the exact location of the gene for 6K on the 26S mRNA, given sufficient amino acid sequencing information. Furthermore, such a nucleotide sequence should allow prediction of the complete amino acid sequence of all of the polypeptides encoded by the polygenic message. A precise identification of the termini of the individual genes cannot be made simply from the nucleotide sequence, however, because the proteins encoded by the 26S mRNA are produced by a series of nascent proteolytic cleavages rather than by individual initiation and termination events. To deduce

exactly where each gene begins and ends, direct amino acid sequencing results are needed. We report here, therefore, the partial amino terminal amino acid sequence of the SFV polypeptides p62, E3, and 6K. In addition, we include a partial NH_2-terminal amino acid sequence of the Sindbis virus homolog of 6K, the 4.2K protein (17).

MATERIALS AND METHODS

Labeling and separation of Sindbis virus and SFV polypeptides. BHK cells were infected with either Sindbis virus or SFV at a multiplicity of 50 to 100 in Dulbecco modified Eagle medium supplemented with 1% calf serum and actinomycin D (2 μg/ml, Calbiochem). The cells were labeled 5 h after infection with either L-[4,5-^{3}H(N)]leucine (New England Nuclear Corp.; 58 Ci/mmol) or L-[4,5-^{3}H(N)]-isoleucine (New England Nuclear Corp.; 100 Ci/mmol). After a 2-h labeling period, the infected cells were solubilized in sodium dodecyl sulfate (SDS)-gel electrophoresis sample buffer at a final concentration of 2% SDS, 1% mercaptoethanol, 15% glycerol, 5 mM sodium phosphate (pH, 7.0), and 0.1 M dithiothreitol and subjected to electrophoresis on a 20% SDS-polyacrylamide gel. To identify the labeled polypeptides by autoradiography, small amounts of infected cells labeled with [^{35}S]methionine (>500 Ci/mmol; Amersham/Searle) were added to each tritium-labeled sample before SDS-polyacrylamide gel electrophoresis.

SDS-polyacrylamide gel electrophoresis. SDS-polyacrylamide gel electrophoresis was performed by using 1-mm-thick, 13.5-cm-long, 20% polyacrylamide slab gels as previously described (16).

Electroelution of the proteins from polyacrylamide gels. The labeled proteins were identified by autoradiography and excised from the gel. The gel slice was separated from the paper backing and placed in Spectrapor 3 dialysis tubing (Spectrum Medical Industries) containing electroelution buffer (0.01 M NH_4HCO_3, 0.1% SDS, 5% mercaptoethanol). The bags were sealed, and the gel slice was hydrated for 30 min. The dialysis bag was then placed between two electrodes, which were approximately 10 cm apart, and elution buffer was added so that the solution just barely covered the bag. The polypeptides were eluted by electrophoresis with a constant current of 200 mA for a period of 3 to 6 h. To recover the protein, the current was reversed briefly, and the elution buffer was removed, filtered through cotton, and lyophilized. The protein was then resuspended in water (1 to 2 ml) and dialyzed against 0.03% SDS at 4°C to remove salts and to reduce the concentration of SDS. Three 2-liter changes of the dialysate were made over a 36-h period. The bag was washed with water, and the protein was lyophilized.

Amino acid sequencing. The lyophilized protein was resuspended in approximately 0.5 ml of water and subjected to sequential Edman degradation on a Beckman 890 automated sequenator. Polybrene was used as a carrier, and a 0.33 M Quadrol program similar to that described by Hunkapiller and Hood was used for each determination (12). The residues from the sequential analyses were dried, transferred into scintillation vials with two 150-μl washes of methanol, and counted in 10 ml of Budget Solve scintillation cocktail (Research Products International).

RESULTS

Partial NH_2-terminal amino acid sequence of SFV 6K and Sindbis virus 4.2K. The SFV 6K protein contained a methionine residue at position 7 and leucine residues at positions 10, 17, and 20 (Fig. 1). The Sindbis virus 4.2K protein also contained a methionine at position 7 and a leucine at position 10 (Fig. 1). Unlike 6K, 4.2K did not contain leucine residues at positions 17 and 20.

Partial NH_2-terminal amino acid sequence of SFV p62 and E3. The NH_2 terminus of SFV p62, the precursor to the envelope glycoproteins E3 and E2, contained a methionine residue at position 8, leucine residues at positions 4 and 12, and an isoleucine residue at position 5 (Fig. 2). The partial NH_2-terminal amino acid sequence of E3 was identical to that of p62 (Fig. 2).

DISCUSSION

Our earlier experiments (17) confirmed that the genes coding for the structural proteins of

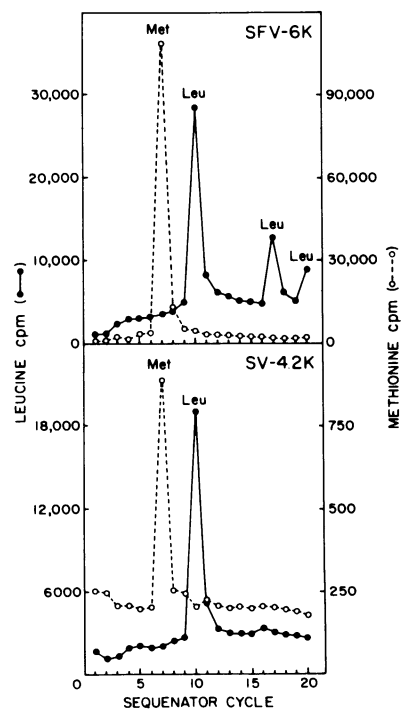

FIG. 1. *Partial NH_2-terminal amino acid sequence of the SFV 6K and Sindbis virus 4.2K proteins. Samples were labeled and prepared for sequential Edman degradation as described in the text.*

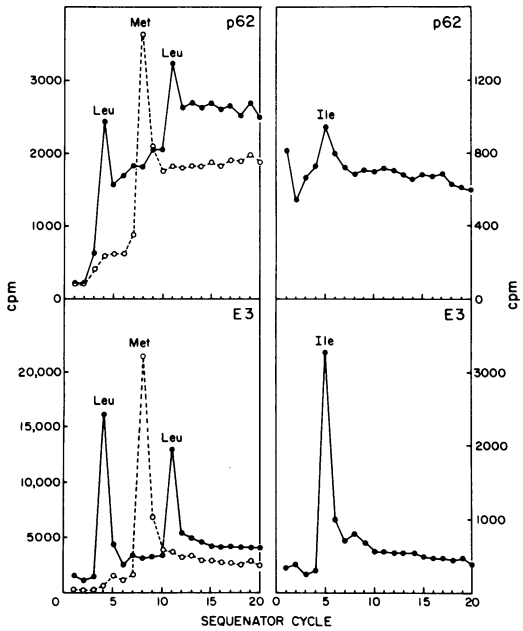

FIG. 2. *Partial NH_2-terminal amino acid sequence of SFV p62 and E3. Samples were labeled and prepared for sequential Edman degradation as described in the text.*

SFV are arranged on the subgenomic 26S mRNA in the order 5′-C-p62-E1-3′ (4). In addition, they showed that the gene coding for the nonstructural 6K protein was located downstream from the gene encoding p62. They did not, however, allow determination of whether the 6K gene was located to the 5′ side or to the 3′ side of the gene for E1. Examination of the predicted amino acid sequence encoded by the 26S mRNA deduced from the sequence of cloned DNA complementary to the 26S mRNAs (6, 7) reveals that the partial amino-terminal amino acid sequence of 6K, which was determined here, occurs only once and is encoded on the mRNA between the regions which encode the COOH terminus of p62 and the NH_2 terminus of E1. This region can code for 60 amino acids or a polypeptide of 6,590 daltons. This is in good agreement with the value of 6,000 daltons we determined for 6K from its migration in SDS gels (17). We conclude, therefore, that the gene for 6K resides between those of p62 and E1 on the SFV 26S mRNA.

The homolog of the SFV 6K protein, Sindbis 4.2K (16), contains a partial NH_2-terminal amino acid sequence similar to, but not identical with, that of 6K. Examination of the predicted amino acid sequence of Sindbis virus 26S mRNA, determined by nucleotide sequencing of cDNA complementary to the 26S RNA (C. M. Rice and J. H. Strauss, Proc. Natl. Acad. Sci. U.S.A., in press), has shown that the gene encoding 4.2K, like the gene for 6K of SFV, is between the genes encoding PE2 and E1. This region can encode 55 amino acids and therefore should yield a protein larger than the 4,200 daltons we ascribed to 4.2K. Given the uncertainties of molecular weight determination by SDS-polyacrylamide gel electrophoresis, we do not feel this discrepancy is significant. It seems clear then, that the gene for 4.2K is located between those for PE2 and E1.

p62, the precursor to E2 and E3 has a partial NH_2-terminal amino acid sequence identical to that of E3. This corroborates the previous observation of Garoff and Soderlund (10), who showed that proteolytic fragments derived from the COOH termini of p62 and E2 were identical. Furthermore, our results indicate that the NH_2 terminus of E3 is not further modified after release from p62.

The order of the genes on the subgenomic 26S mRNA of both Sindbis virus and SFV is now complete. The gene order of SFV is 5′-C-E3-E2-6K-E1-3′. Similarly, the Sindbis virus genes are arranged 5′-C-9.8K-E2-4.2K-E1-3′.

We previously suggested that 6K of SFV and 4.2K of Sindbis virus functioned in some manner to aid the folding of the viral glycoproteins as they are inserted into the membrane of the endoplasmic reticulum (17). Knowledge of the location of the genes for these peptides makes speculation as to the actual role that they play now possible. For simplicity, the discussion will be limited to the 6K protein of SFV and its role in the synthesis and insertion of E1 into the membrane of the endoplasmic reticulum. The two models discussed here apply equally well, however, to the role of the 4.2K protein of Sindbis virus.

The bulk of the polypeptide chains of both newly synthesized E1 and p62 are sequestered in the RER. Proteolytic digestion of vesicles derived from the RER has no effect on E1 and removes no more than 3,000 daltons from p62 (10). Peptide mapping has shown that it is the COOH terminus of p62 which is susceptible to proteolysis in such vesicles (10). This suggests that p62 spans the membrane of the RER and that only its COOH terminus is exposed on the cytoplasmic face of the RER. The resistance of E1 to digestion suggests that little, if any, of it is present in the cytoplasm. Each ribosome which participates in the translation of the 26S mRNA synthesizes in sequence C, p62, 6K, and E1 and terminates only at the 3′ end of the mRNA. Mechanisms must exist to allow, first, the deposition of p62 in the membrane of the RER as a membrane-spanning polypeptide and, subse-

quently, the insertion of E1 into the RER so that it is not susceptible to proteolytic digestion. Bonatti and Blobel (3) have suggested that the NH_2 terminus of p62 serves as a signal peptide to initiate binding of p62 to the RER. We suggest that the function of 6K is to cause both the translocation of E1 into the RER and the exposure of the COOH terminus of p62 in the cytoplasm.

There are at least two ways in which this process could occur (Fig. 3). First, 6K could serve the same function as the signal peptides (1, 2) which exist at the NH_2 termini of the nascent forms of many secreted and membrane-associated proteins (Fig. 3, scheme a). For this to occur the cleavage which separates p62 and 6K would have to take place in the cytoplasm very soon after the synthesis of the cleavage site between the two proteins. Once separated from p62, 6K could serve to transfer the NH_2 terminus of E1 across the membrane into the lumen of the RER. 6K would then presumably be removed by a signal peptidase present in the lumen of the RER. Alternatively, the cleavage between p62 and 6K could be delayed, and the nascent chain could loop back into the membrane with 6K serving as a spacer between the COOH terminus of p62 in the cytoplasm and the NH_2 terminus of E1 in the lumen (Fig. 4, scheme b). Excision of 6K would then yield both glycoproteins in the proper orientation.

For several reasons, we favor the second model. Although 6K resembles a signal peptide in that it is predominantly hydrophobic in composition, it differs from typical signal peptides both in that it consists of 60 amino acids, rather than the usual 20 to 25, and in that it is sufficiently stable metabolically to be readily detected in vivo. Many bonafide signal peptides appear to be very unstable in vivo (11). Additionally, the lag with which labeled p62 and PE2 appear after the onset of labeling of infected cells with radioactive amino acids lends support to the second model. Scheme b predicts that the release of newly synthesized p62 or PE2 from the nascent chain is delayed significantly and occurs only after the subsequent translation of the 6K gene and a portion of the E1 gene has occurred. If this is the case, alphavirus-infected cells which have been labeled with radioactive amino acids for a period of time shorter than the time it takes for a ribosome to synthesize 6K and a portion of E1 will not contain any labeled p62 or PE2 which has yet become a discrete species. Instead, all of the labeled p62 or PE2 in such cells will be present in heterogeneous nascent chains and thus be unrecognizable as p62 or PE2 by SDS-polyacrylamide gel electrophoresis. There is some reason to believe that this is the case. We found several years ago that Sindbis virus-infected chicken cells which have been labeled for 60 s with [^{35}S]methionine contain labeled C and E1, but no labeled PE2 (14). Very similar results are obtained when SFV-infected BHK cells labeled with [^{35}S]methionine for 60 s are analyzed (W. Welch and B. Sefton, unpublished data). Although there are clearly several interpretations for the peculiar lag with which p62 and PE2 become labeled, this phenomenon is fully consistent with the model in which the excision of 6K from the nascent chain is delayed significantly relative to its synthesis.

Our second model is somewhat analogous to the mechanism which has been proposed to

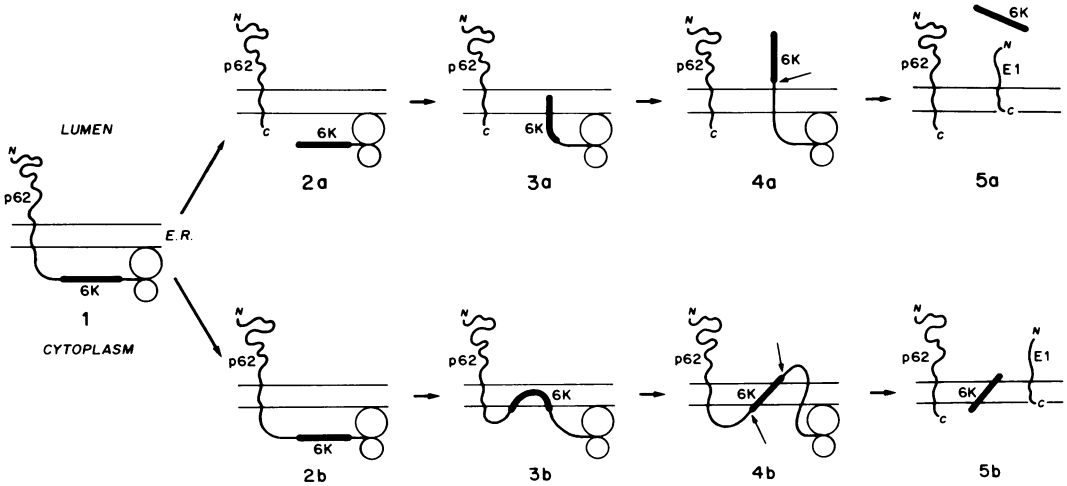

FIG. 3. *Two possible mechanisms for the insertion of the E1 glycoprotein into the endoplasmic reticulum. The details of these two models are discussed in the text. For clarity, the 26S mRNA is not pictured here.*

explain the transfer of ovalbumin across the RER membrane (13). Unlike many secreted proteins nascent ovalbumin does not contain an amino-terminal signal peptide which is removed as the protein is transferred across the membrane. It has been suggested that the protein contains instead a signal-like sequence in the middle of the ovalbumin polypeptide chain which leads the nascent ovalbumin protein across the membrane into the lumen of the RER and which is not excised from the protein. The SFV nascent polypeptide chain would differ from ovalbumin in that the internal membrane insertion peptide, 6K, is eventually excised.

The 4.2K protein of Sindbis virus and the 6K protein of SFV appear to be homologous in almost all respects. We originally termed the peptide encoded by Sindbis virus 4.2K because of its apparent mobility on SDS gels (16). Since this peptide probably contains 55 amino acids (Rice and Strauss, in press), this name now may be somewhat misleading. We suggest, in the absence of a more functional, nonnumerical name, that it will be simplest to refer to both of the small late nonstructural peptides of SFV and Sindbis virus as 6K.

ACKNOWLEDGMENTS

This work was supported by grant PCM 77-25974 from the National Science Foundation and Public Health Service grants CA-14195 and CA-17289 from the National Institutes of Health.

LITERATURE CITED

1. **Blobel, G., and B. Dobberstein.** 1975. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobin light chains on membrane-bound ribosomes of murine myeloma. J. Cell Biol. **67**:835–851.
2. **Blobel, G., P. Walter, C. N. Chang, B. Goldman, A. H. Erickson, and V. K. Lingappa.** 1979. Translocation of proteins across membranes: the signal hypothesis and beyond. Symp. Soc. Exp. Biol. **33**:9–36.
3. **Bonatti, S., and G. Blobel.** 1979. Absence of a cleavable signal sequence in Sindbis virus glycoprotein PE2. J. Biol. Chem. **254**:12261–12264.
4. **Clegg, J. C. S.** 1975. Sequential translation of capsid and membrane protein genes in arbovirus-infected cells. Nature (London) **254**:454–455.
5. **Clegg, J. C. S., and S. I. T. Kennedy.** 1975. Initiation of synthesis of the structural proteins of Semliki Forest virus. J. Mol. Biol. **97**:401–411.
6. **Garoff, H., A. M. Frischauf, K. Simons, H. Lehrach, and H. Delius.** 1980. Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins. Nature (London) **288**:236–241.
7. **Garoff, H., A.-M. Frischauf, K. Simons, H. Lehrach, and H. Delius.** 1980. The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino terminal region. Proc. Natl. Acad. Sci. U.S.A. **77**:6376–6380.
8. **Garoff, H., K. Simons, and B. Dobberstein.** 1978. Assembly of the Semliki Forest virus membrane glycoproteins in the membrane of the endoplasmic reticulum *in vitro*. J. Mol. Biol. **124**:587–600.
9. **Garoff, H., K. Simons, and O. Renkonen.** 1974. Isolation and characterization of the membrane proteins of Semliki Forest virus. Virology **61**:493–504.
10. **Garoff, H., and H. Soderlund.** 1978. The amphiphilic membrane glycoproteins of Semliki Forest virus are attached to the lipid bilayer by their COOH-terminal ends. J. Mol. Biol. **124**:535–549.
11. **Habener, J. F., M. Rosenblatt, P. C. Dee, and J. T. Potts.** 1979. Cellular processing of pre-proparathyroid hormone involves rapid hydrolysis of the leader sequence. J. Biol. Chem. **254**:10596–10599.
12. **Hunkapiller, M. W., and L. E. Hood.** 1978. Direct microsequence analysis of polypeptides using an improved sequenator, a nonprotein carrier (polybrene) and high pressure liquid chromatography. Biochemistry **17**:2124–2133.
13. **Lingappa, V. R., J. R. Lingappa, and G. Blobel.** 1979. Chicken ovalbumin contains an internal signal sequence. Nature (London) **281**:117–121.
14. **Sefton, B. M.** 1977. Immediate glycosylation of Sindbis virus membrane proteins. Cell **10**:659–668.
15. **Simons, K., S. Keränen, and L. Kääriäinen.** 1973. Identification of a precursor for one of the Semliki Forest virus membrane proteins. FEBS Lett. **29**:87–91.
16. **Welch, W. J., and B. M. Sefton.** 1979. Two small virus-specific polypeptides are produced during infection with Sindbis virus. J. Virol. **29**:1186–1195.
17. **Welch, W. J., and B. M. Sefton.** 1980. Characterization of a small nonstructural viral polypeptide present late during infection of BHK cells by Semliki Forest virus. J. Virol. **33**:230–237.