Proc. Natl. Acad. Sci. USA
Vol. 75, No. 6, pp. 2722–2726, June 1978
Biochemistry

Amino-terminal sequence analysis of the structural proteins of Sindbis virus

(membrane glycoproteins/togaviruses/microsequencing)

JOHN R. BELL, MICHAEL W. HUNKAPILLER, LEROY E. HOOD, AND JAMES H. STRAUSS

Division of Biology, California Institute of Technology, Pasadena, California 91125

Communicated by N. H. Horowitz, April 7, 1978

ABSTRACT **The structural proteins of Sindbis virus, an enveloped virus which belongs to the Togavirus family, have been subjected to automated Edman degradation using improved techniques. Extensive NH_2-terminal sequences of about 50 residues were determined for each of the two membrane glycoproteins. In both cases the NH_2 terminus of the molecule was found to be similar in composition to typical water-soluble proteins. The viral capsid protein was found to have a blocked α-amino group. This is consistent with other observations that viral proteins derived from the NH_2 terminus of precursor molecules are often blocked.**

Sindbis virus is a simple enveloped virus which matures by budding through a modified host cell plasma membrane (1). The virus particle contains only three structural proteins (2). One of these, the capsid protein, can be isolated from virus particles in the form of a nucleocapsid structure also containing the viral RNA (3). In the intact virus particle, this nucleocapsid is surrounded by a lipid envelope, largely in the form of a bilayer (4). Associated with the envelope are the other two structural proteins, glycoproteins E1 and E2. Although most of their mass is external to the lipid bilayer, they have the properties of integral membrane proteins. Mild detergent treatment is required to separate E1 and E2 from the other components of the virus (3, 5), and at least one is a transmembrane protein because it extends completely through the bilayer and can be crosslinked to the capsid protein by suitable reagents (6).

It has been shown that all three Sindbis structural proteins as well as those of the closely related Semliki Forest virus are formed by cleavage of a single precursor polypeptide (7–9), probably while the polypeptide chain is still nascent. The capsid protein is released into the cytoplasm of the cell; in contrast, E1 and E2 are inserted into the lumen of the rough endoplasmic reticulum (10). At some time after synthesis, they appear glycosylated at the cell plasmalemma, where they diffuse freely over the surface of the cell (11). The budding process appears to involve an interaction between preformed nucleocapsids and that region of one or both of the envelope proteins exposed at the cytoplasmic face of the plasmalemma. This binding of capsid protein to a region of the envelope protein is specific enough that host cell membrane proteins are excluded from the bilayer of the budding virus and are not found in mature virus particles (12).

Because of the relative ease of purifying material for study and the availability of mutants containing defects in the viral structural proteins, Sindbis virus offers an attractive system for the study of integral membrane proteins and their interactions with membranes and with other proteins. Recent advances in the techniques of automated protein sequencing have made practical the sequencing of the proteins of viruses, such as Sindbis, which are difficult to obtain in large quantities. We have begun sequence studies on the Sindbis virus structural proteins, using microsequencing techniques developed by two of the authors (13), with the aim of explaining the functional roles of the viral proteins on the basis of their sequence and structure.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Protein Preparation. The growth and purification of Sindbis virus and the separation of the structural proteins will be described in detail elsewhere. Briefly, the HR strain of Sindbis virus, grown at low multiplicity of infection in primary chicken embryo fibroblasts, was concentrated from culture fluid by precipitation with polyethylene glycol followed by pelleting onto a fluorocarbon cushion. The resuspended virus was then purified by sucrose gradient velocity centrifugation followed by isopycnic centrifugation in sucrose/^2H_2O gradients. These steps of the procedure were similar to those described previously (14). After disruption with Triton X-100, capsids were removed by centrifugation and the two envelope glycoproteins were separated by passage through a column of glass wool. In Triton X-100, E2 binds to glass and is retained on the column while, under appropriate conditions, E1 is not retained. E2 was recovered by washing the column with a high-salt high-pH Triton X-100 buffer. Both E1 and E2 were treated with sodium dodecyl sulfate and concentrated by Amicon filtration and precipitation with ethanol. The capsid protein was freed of viral RNA by hydroxylapatite chromatography after solubilization with sodium dodecyl sulfate.

Operation of the Spinning Cup Sequenator. The technique and instrumentation used for automated Edman degradation in the spinning cup sequenator is described in detail elsewhere (13). A Beckman Instruments spinning cup sequenator was extensively modified to provide an improved vacuum system, and solvents and reagents were delivered through a system of specially constructed zero-dead-volume valves. Twenty to 50 nmol of purified protein was loaded onto the instrument in anhydrous trifluoroacetic acid and the instrument was run under an essentially standard Quadrol protein program with double cleavage (13). Polybrene was used as a carrier to retain the protein in the spinning cup.

The sequenator also was modified to perform the automated conversion of the anilinothiazolinones to the corresponding amino acid phenylthiohydantoins by treatment with aqueous trifluoroacetic acid. The reproducibility of this conversion results in stable background levels of amino acid derivatives. Amino acid phenylthiohydantoins were identified and quantitated by reverse-phase high-pressure liquid chromatography on Du Pont Zorbax ODS.

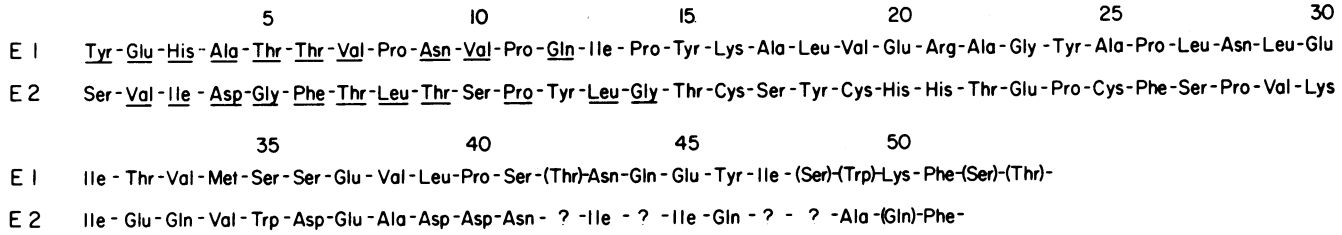

FIG. 1. NH_2-Terminal amino acid sequences of the Sindbis virus envelope proteins, E1 and E2. (), Some uncertainty in assignment; ?, no assignment. Underlined amino acids were also identified in preliminary runs on the solid-phase sequenator.

Solid Phase Sequenator. A new type of Edman sequenator using gas phase Edman reagents reacting with proteins on a solid phase support will be described elsewhere. One to 2 nmol of protein was dissolved in formic acid and the solution was dried onto underivatized porous glass. The glass support with adsorbed protein was loaded into a tubular cartridge in the sequenator, where gas phase reagents were passed through the cartridge to effect the Edman degradation. ^{35}S-Labeled phenylisothiocyanate was used as a reagent to introduce radioactive label into the released amino acid derivatives. Amino acid thiazolinones released by the degradation were washed from the porous glass with a nonpolar solvent and collected in fraction collector tubes. Thiazolinones were automatically converted to amino acid phenylthiohydantoins by treatment with aqueous trifluoroacetic acid delivered by the sequenator to the fraction collector tubes. The amino acid phenylthiohydantoins were identified by thin-layer chromatography (15) followed by autoradiography to detect the ^{35}S-labeled derivatives.

RESULTS

In preliminary experiments, less than 2 nmol of each of the Sindbis envelope proteins was sequenced on the solid-phase sequenator and more than 10 residues were identified in each case. Amino acids identified in these experiments are underlined in Fig. 1. These results were confirmed and extended with the analysis of independent preparations of the envelope proteins on the spinning cup sequenator.

Twenty-five to 50 nmol of purified E1 and E2 were sequenced on the spinning cup sequenator. The NH_2-terminal sequences for the two proteins are shown in Fig. 1, and yields of the various amino acids, as determined by liquid chromatography of the phenylthiohydantoins, for each sequenator cycle are shown in Figs. 2 and 3 for E1 and E2, respectively. In both cases, approximately 50 positions in the sequence could be determined. A major factor in the sequence determination toward the ends of the runs is the stability of the background level of amino acid phenylthiohydantoins found at each step. As can be seen in Figs. 2 and 3, in some cases amino acid derivatives that were produced at a level of 15% above the background could be identified in the sequence. In the case of E2 (Fig. 3), the lag increased (and repetitive yields decreased slightly) near cycle 15. This may be due to interchain disulfide bond formation between the cysteines at positions 16 and 19. Such bonding might be expected to restrict the accessibility of an exposed NH_2-terminal amino group at this region to the coupling reagent, the result being a temporary decrease in the efficiency of the coupling reaction. In both cases, amino acids at each position at least through position 40 could be positively identified. The fact that neither obvious blanks in the sequence nor anomalous peaks on the chromatograms were encountered indicates that there are no attached carbohydrate chains over the first 40 residues of either membrane glycoprotein.

The amino acids identified in the NH_2-terminal sequences of E1 and E2 were grouped into several classes on the basis of the character of their side chains, and relative abundance of each of these groups of amino acids, on a mol % basis, are presented in Table 1. Also shown in the table are the compositions of the complete proteins, with the amino acids grouped in the same way. Neither the complete proteins nor the NH_2-terminal 50 residues of E1 and E2 contain an excess of hydrophobic amino acids. Compared to the proteins as a whole, the NH_2-terminal regions of both E1 and E2 tend to contain relatively more nonbasic than basic polar amino acids and more aromatic than aliphatic nonpolar residues. In addition, proline is somewhat enriched in the NH_2-terminal region of E1.

The capsid protein was also analyzed on the spinning cup sequenator and was found to contain a blocked NH_2-terminal amino group. When 20 nmol of purified protein was loaded, less than 0.2 nmol above background of any one amino acid derivative could be detected in the first seven sequenator cycles, although the background amount of amino acids increased in a manner similar to that seen for E1 and E2.

DISCUSSION

Integral membrane proteins generally are considered to share certain common features, as described by Singer (16). A region of the molecule is presumed to be buried in the hydrophobic interior of the membrane, and the amino acid sequence of this region of the protein should reflect its hydrophobic character.

Table 1. Composition of Sindbis glycoproteins

Amino acid side chain type	E1		E2	
	NH_2-terminal region	Complete protein	NH_2-terminal region	Complete protein
Basic polar (His, Lys, Arg)	8	10	6	15
Nonbasic polar (Asx, Glx, Ser, Thr)	38	33	40	31
Aliphatic nonpolar (Gly, Ala, Met, Val, Ile, Leu)	34	38	28	35
Aromatic nonpolar (Tyr, Phe, Trp)	11	9	13	9
Pro	9	6	6	7
Cys	0	3	6	3

The relative occurrence of each group of amino acids is expressed on a mol % basis. Data for the NH_2-terminal regions of the proteins were computed from Fig. 1, including all indicated residues.

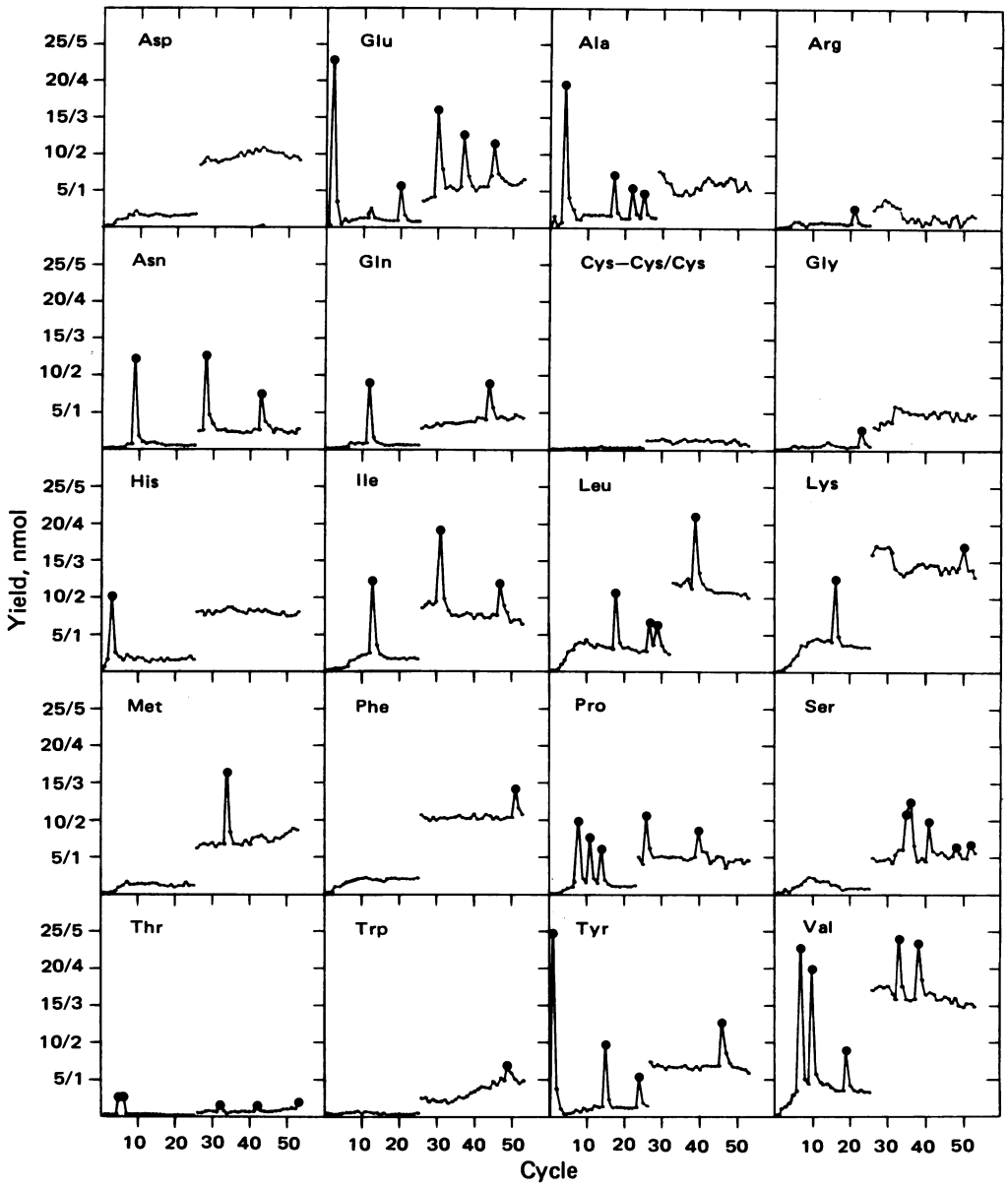

FIG. 2. Yields of amino acid phenylthiohydantoins from an NH_2-terminal spinning cup sequenator analysis of 50 nmol of Sindbis glycoprotein E1 (peptide molecular weight, ~45,000). Aliquots of each cycle were analyzed by high-pressure liquid chromatography, peaks were quantitated by comparison with a standard mixture of amino acid phenylthiohydantoins, and the yields were normalized to an injection of 100% of the sample. The ordinate gives two scales which differ 5-fold in sensitivity. Early cycles are plotted on the left (less sensitive) scale, and later cycles are plotted on the right (more sensitive) scale.

However, the regions of the protein that extend beyond the membrane into an aqueous environment should be folded so as to expose only hydrophilic groups to the environment, in a manner similar to that of typical water-soluble proteins. The available sequence information on a few integral membrane proteins tends to confirm this arrangement. For example, glycophorin, a major erythrocyte membrane component, contains a stretch of 23 amino acids, only 3 of which (2 threonines and 1 serine) could be considered hydrophilic (17). No charged amino acids are found in this region. On either side of this region of the molecule, which is protected by the membrane from digestion by protease, the sequence contains a more typical distribution of polar amino acids.

Both of the Sindbis glycoproteins appear to also contain a hydrophobic region. Upon treatment of the closely related Semliki Forest virus with low concentration of a mild detergent (5), the envelope protein and lipid are released from the nucleocapsid in the form of lipoprotein complexes. At higher detergent concentrations, the lipid is displaced from the protein and replaced with detergent. Furthermore, in the case of both Sindbis virus (18) and Semliki Forest virus (19), protease treatment extensively degrades the envelope glycoproteins but leaves attached to the treated particle small polypeptides derived from these proteins. These are quite hydrophobic in character and consist primarily of those regions of the proteins buried in the lipid bilayer.

The NH_2-terminal 50 residues of E1 and E2 contain a proportion of polar amino acids similar to that found in typical water-soluble proteins (20). Therefore, this region of each protein is probably exposed to an aqueous environment rather than interacting directly with the envelope lipids. Further, the NH_2-terminal region can be expected to be external to the viral envelope by comparison with other viral (21) and nonviral (17, 22) membrane proteins. Although the carbohydrate of E1 and

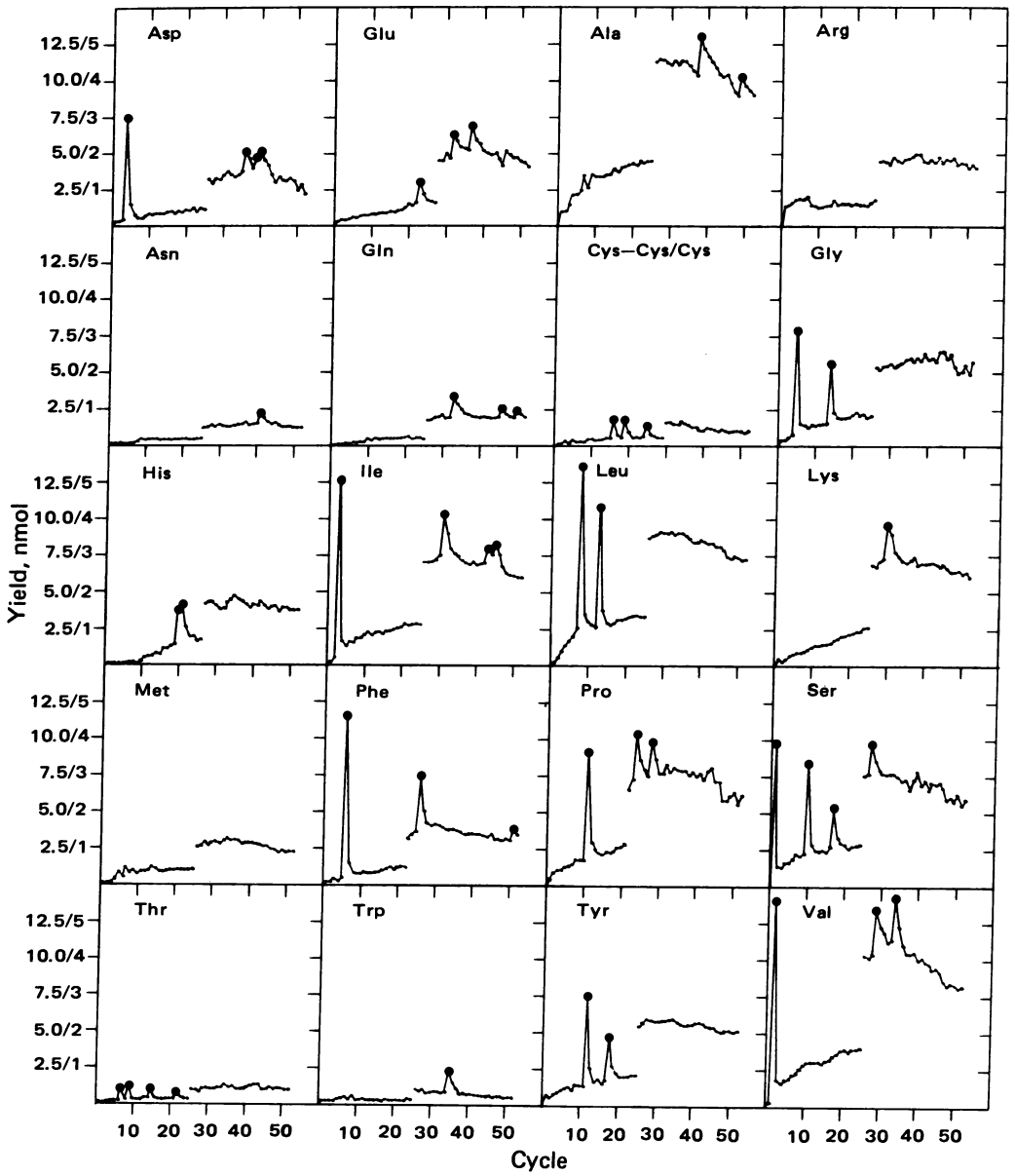

FIG. 3. Yields of amino acid phenylthiohydantoins from an NH_2-terminal spinning cup sequenator analysis of 25 nmol of Sindbis glycoprotein E2 (peptide molecular weight, ~48,000). Yields were calculated as described in the legend to Fig. 2, and the data are plotted in a similar fashion with a 2.5-fold expansion in scale for the later cycles.

E2 is known to be external to the viral envelope, no evidence of carbohydrate attachment is seen at least through the first 40 residues of each protein. This is in contrast to the situation found in glycophorin (15) in which carbohydrate is attached to the second residue and there are 16 carbohydrate attachment sites in the first 50 residues.

In many virus systems, the production of functional structural proteins involves proteolytic cleavage (23). In the case of Sindbis virus, one large precursor is cleaved, probably by host cell enzymes, to produce all three structural proteins (7–9). The capsid protein is located at the NH_2 terminus of this precursor, although it is not known if its sequence extends completely to the precursor NH_2 terminus. We have found that the NH_2 terminus of the capsid protein is blocked, raising the possibility that this same modified residue is the NH_2-terminal residue of the precursor. Other viruses are also known to produce blocked structural proteins from the NH_2-terminal ends of precursor molecules. Mengo virus is similar to Sindbis virus in that only the δ protein of the virus particle is blocked and it is derived from the NH_2 terminus of a precursor to all the structural proteins (24). In the case of Sendai virus, there are several different mRNAs for the structural proteins. However, proteolytic cleavage still occurs in the formation of infectious virus particles. The blocked protein, F_0, is cleaved and the two products, joined by disulfide bonds, form the functional F protein of the virus (25). In addition, the NH_2-terminal residue of the adenovirus 2 hexon protein is acetylated (26). It may be that viral structural proteins and their precursor polypeptides are often blocked, perhaps to increase the resistance of these molecules to degradation by exopeptidases. In all cases of which we are aware, when a blocked viral protein is formed by cleavage of a precursor, the blocked product is located at the NH_2 terminus of the precursor polypeptide. This would seem to indicate that, in these cases, the blocking group is added during or shortly after translation but before extensive proteolytic processing takes place.

The identification of a free NH_2-terminal lysine for the capsid protein of Semliki Forest virus has been reported (27). However, nonquantitative methods (the dansyl technique) were used, and lysine is the most prevalent amino acid in the Semliki Forest virus capsid protein. Because the two viruses are closely related, it would be surprising if the NH_2 terminus of the Semliki Forest virus capsid protein is not also blocked.

The NH_2 termini of E1 and E2 are both created by proteolytic cleavage, raising the possibility that they might share some common structural features reflecting similarities at the two cleavage sites. However, no similarities are apparent, other than the fact that they both contain a reasonable number of polar amino acids near the NH_2 terminus. Sequence data are also available for some other viral proteins whose NH_2 termini are generated by cleavage events, including the glycoprotein HA2 of influenza (21) and nonglycosylated proteins such as the capsid proteins α, β, and γ of Mengo virus (24) and p30 of the mammalian RNA leukemia viruses (28). The NH_2 termini of these proteins show little apparent sequence similarity with each other or with E1 or E2 of Sindbis virus. For example, the NH_2-terminal 10 residues of HA2 of influenza (which are highly conserved in all strains studied) are all hydrophobic (21). Furthermore, residues 3–9 form a palindrome, whereas E1 and E2 of Sindbis virus and the other proteins mentioned above do not contain such a sequence. Despite the fact that they exhibit little similarity from one protein to another, the NH_2 termini of some of these molecules have been shown to be highly conserved, so the particular sequence must be an important factor in the delineation of the cleavage site or in the structure or function of the protein product. The most striking example of this sequence conservation is provided by p30 of the mammalian RNA leukemia viruses. Viruses from the cat, mouse, rat, gibbon ape, and baboon have been studied (28), and the p30 protein of each begins with the sequence Pro-Leu-Arg-.

An extensive amount of sequence data has been obtained for the Sindbis virus glycoproteins from single sequenator runs on protein preparations of 50 nmol or less. Even so, we note from Figs. 2 and 3 that the limiting factor in the number of residues that could be identified is the amount of amino acid derivative released in a sequenator cycle above the background level of that derivative. Although the levels of derivatives found in late cycles are on the order of 0.5–2 nmol, the instrument detection limit for the various derivatives is <10 pmol. Thus, even smaller amounts (e.g., 1 nmol) of protein would have given 30 or more residues because the background level is proportional to the amount of protein loaded. In test runs using the same sequencing methodology, Hunkapillar and Hood (13) were able to identify the NH_2-terminal 47 residues of myoglobin using only 0.2 nmol of protein. It is also readily apparent that these methods can be applied to isolated fragments of these proteins such as the fragments produced by CNBr cleavage, at either methionine or tryptophan residues, or by partial proteolytic digestion, and that such a strategy can be used to obtain the primary sequence of these proteins from quite small amounts of material.

The preliminary sequence data obtained by using the solid-phase instrument were obtained by Dr. W. J. Dreyer and his colleagues and we express our gratitude to him for this work. Edith M. Lenches and Mary S. Martin provided expert technical assistance in the growth of the virus and its purification. This work was supported by Grants GM 06965 and AI110793 from the National Institutes of Health and by Grant PCM 77-26728 from the National Science Foundation.

1. Strauss, J. H. & Strauss, E. G. (1977) in *The Molecular Biology of Animal Viruses*, ed. Nayak, D. P. (Marcel Dekker, New York), Vol. 1, pp. 111–166.
2. Schlesinger, M. J., Schlesinger, S. S. & Burge, B. W. (1972) *Virology* **47**, 539–541.
3. Strauss, J. H., Burge, B. W., Pfefferkorn, E. R. & Darnell, J. E., Jr. (1968) *Proc. Natl. Acad. Sci. USA* **59**, 533–537.
4. Harrison, S. C., David, A., Jumblatt, J. & Darnell, J. E., Jr. (1971) *J. Mol. Biol.* **60**, 523–528.
5. Helenius, A. & Söderlund, H. (1973) *Biochim. Biophys. Acta* **307**, 287–300.
6. Garoff, H. & Simons, K. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 3988–3992.
7. Schlesinger, M. J. & Schlesinger, S. (1973) *J. Virol.* **11**, 1013–1016.
8. Simmons, D. T. & Strauss, J. H. (1974) *J. Mol. Biol.* **86**, 397–409.
9. Clegg, J. C. S. & Kennedy, S. I. T. (1975) *J. Mol. Biol.* **97**, 401–411.
10. Wirth, D. F., Katz, F., Small, B. & Lodish, H. F. (1977) *Cell* **10**, 253–263.
11. Birdwell, C. R. & Strauss, J. H. (1974) *J. Virol.* **14**, 366–374.
12. Pfefferkorn, E. R. & Clifford, R. L. (1964) *Virology* **23**, 217–223.
13. Hunkapiller, M. W. & Hood, L. E. (1978) *Biochemistry*, in press.
14. Strauss, J. H., Burge, B. W. & Darnell, J. E. (1969) *Virology* **37**, 367–376.
15. Edman, P. & Sjöquist, J. (1956) *Acta Chem. Scand.* **10**, 1507–1509.
16. Singer, S. J. (1971) in *Structure and Function of Biological Membranes*, ed. Rothfield, L. I. (Academic, New York), pp. 145–222.
17. Tomita, M. & Marchesi, V. T. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2964–2968.
18. Compans, R. W. (1971) *Nature New Biol.* **229**, 114–116.
19. Utermann, G. & Simons, K. (1974) *J. Mol. Biol.* **85**, 569–587.
20. Capaldi, R. A. & Vanderkovi, G. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 930–932.
21. Skehel, J. J. & Waterfield, M. D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 93–97.
22. Ozols, J., Gerard, C. & Nobrega, F. G. (1976) *J. Biol. Chem.* **251**, 6767–6774.
23. Shapiro, S. Z. & August, J. T. (1976) *Biochim. Biophys. Acta* **458**, 375–396.
24. Ziola, B. R. & Scraba, D. G. (1976) *Virology* **71**, 111–121.
25. Scheid, A. & Choppin, P. W. (1977) *Virology* **80**, 54–66.
26. Jörnwall, H., Ohlsson, H. & Phillipson, L. (1974) *Biochem. Biophys. Res. Commun.* **56**, 304–310.
27. Kennedy, S. I. T. & Burke, D. C. (1972) *J. Gen. Virol.* **14**, 87–98.
28. Oroszlan, S., Copeland, T., Summers, M. R., Smythers, G. & Gilden, R. V. (1975) *J. Biol. Chem.* **250**, 6232–6239.