STUDIES ON THE STRUCTURAL PROTEINS OF SINDBIS VIRUS

Thesis by

Charles Moen Rice, III

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1981

(Submitted May 20, 1981)

Acknowledgements

Many people have either directly or indirectly been of great help during the course of this work. Dennis Barrett suggested that Caltech might be a nice place to go to graduate school. My parents provided support and encouragement. Jim and Ellen Strauss critically evaluated this work during and after its completion, and I am indebted to them for their advice, patience, and friendship. The other members of the Strauss group, John Bell, Jeff Mayne, and James Ou were instrumental in collaborative efforts and in the formation of many of the ideas presented in this thesis. Edith Lenches saved me from the tedium of preparing large quantities of virus and hacking up chicken embryos. Lee Hood and Tom Maniatis generously lent us equipment and along with the other members of my committee, Norm Davidson, Jim Hudspeth, Herschel Mitchell, and Ray Owen (honorary member) had many helpful suggestions. Henry Huang taught me the chemical sequencing method (duly acknowledged, but of dubious merit) and was the source of many enjoyable discussions.

Somehow, during this time, Jeff Mayne and the ghost of Terry Frey have managed to convert me into a sports fan. Without volleyball, JRA II, JAP Memorial, Blacker House, and J. A. Prufrock, I might have finished a couple of years ago. David Asai, John and Barb Frelinger (alias the Stanford "gang of three") provided moral support and were always ready to investigate new ways of avoiding work.

Charles Brokaw, Lody Kempees, and the rest of the Caltech administration somehow managed to get all of the bills paid. Funds came from the Earle C. Anthony and Jean Weigle Funds, the Institute, the NIH predoctoral training grant, and probably some other sources. The division cryptographers were spectacular, in particular Connie Katz managed to perform a miracle in typing this thesis. People of such competence and dedication are rare.

Lastly, without the friendship of Toni Claudio and an occasional bottle of wine, none of this would have been possible.

Abstract

Conditions are described for the synthesis of long cDNA transcripts of Sindbis virus 26S and 49S RNA in high yield. This single-stranded cDNA could be cut with Type II restriction endonucleases including Hae III, Hha I, Rsa I, or Taq I to give reproducible patterns of discrete, virus-specific fragments which were suitable for subsequent end-labeling and direct sequence analysis. Using these methods, the strategy used for obtaining nearly the entire 26S RNA sequence from cDNA synthesized in vitro is presented. The 26S RNA is approximately 4.2 kb in length, and from the AUG codon initiating synthesis of the capsid protein, contains an open reading frame for 3735 nucleotides. From this sequence, the amino acid sequences of the encoded virus structural proteins, which include a basic capsid protein and two integral membrane glycoproteins (E1 and E2), as well as the sequences of two nonstructural polypeptides have been deduced. Features of the primary structure of these proteins and the proteolytic cleavage sites involved in their processing are discussed.

The orientation of the virion glycoproteins with respect to the lipid bilayer was studied by digesting intact Sindbis virus with α-chymotrypsin. A single membrane-associated peptide is produced from each of the two virion glycoproteins. These peptides contain covalently attached palmitic acid, are rich in hydrophobic amino acids and are located at the extreme COOH-terminal end of each glycoprotein. Both peptides contain uninterrupted sequences of uncharged amino acids of sufficient length to span the lipid bilayer, and it is suggested that they serve to anchor the viral glycoproteins in the membrane. The properties of these and other well-characterized transmembrane segments are discussed.

Specific antisera to each of the virus structural proteins was produced and used to study the association of the virion glycoproteins and their precursors. E1 and E2 could be cross-linked into heterodimers using bifunctional amino-reactive imidates.

This association is present both in intact virions and infected cells and is stable after solubilization of the virion envelope by Triton X-100. Cross-linking data of pulse-labeled monolayers and cells infected with ts mutants are summarized. These data suggest that PE2 (the precursor to E2) and E2 are in different conformations with respect to E1, and that the glycoprotein precursors synthesized at elevated temperatures have an increased tendency to undergo intracellular aggregation.

Table of Contents

CHAPTER 1

Structure and maturation of Sindbis virus

The purpose of this chapter is to provide an introduction to the infection cycle of alphaviruses in vertebrate cells with particular emphasis on the translation and processing of the virion structural proteins. Several recent reviews have been published (Strauss and Strauss, 1977; Kääriäinen and Söderlund, 1978; Simons et al., 1978; see also, The Togaviruses, Schlesinger, R. W., ed., Academic Press, New York) which summarize in detail what is known about the molecular biology of alphaviruses.

Sindbis virus, and the closely related Semliki Forest virus, are the most extensively studied members in the alphavirus genus of the togavirus family. These simple enveloped viruses are transmitted in nature by blood-sucking arthropods, and infect a wide variety of vertebrate hosts including birds and mammals. The mature virion contains a single 49S genomic RNA complexed with approximately 240 capsid protein (C) (MW ∿30,000 daltons) subunits to form an icosahedral nucleocapsid. This structure is surrounded by a lipid bilayer of host cell origin containing two virus-specific integral membrane glycoproteins (E1 and E2) (each of MW ∿50,000 daltons) organized in spikelike projections on the external surface of the virus particle. These projections, which include the majority of the glycoprotein mass, can be removed by proteolysis of the virion (Compans, 1971; see Chapter 4). The virion envelope is acquired as the nucleocapsid buds through the host cell plasma membrane during the final stages of virus maturation (reviewed by Murphy, 1980).

The infection cycle begins with the adsorption of the virus to specific receptors in the host cell plasma membrane (Birdwell and Strauss, 1974). In the case of Semliki Forest virus, these receptors have recently been identified as the HLA-A and HLA-B antigens on human cells and the H2-K and H2-D antigens on mouse cells (Helenius et al., 1978). Following adsorption to the cell surface, the virus could enter the cell either by fusion with the plasma membrane or endocytosis. Fusion of the virion envelope

with the plasma membrane would presumably lead to the incorporation of viral glycoproteins into the host cell surface. This possibility has been tested by using virus-specific antisera in a complement mediated cytotoxicity assay (Fan and Sefton, 1978). Cells infected with Sindbis virus were not sensitive to lysis during the early stages after absorption, suggesting that penetration may not be mediated by fusion with the plasma membrane, in contrast to paramyxoviruses such as Sendai virus (Fan and Sefton, 1978). Support for the second pathway, receptor-mediated endocytosis, has been obtained for Semliki Forest virus by Helenius and coworkers (Helenius et al., 1980). Electron microscopic investigations showed that virus is internalized by endocytosis in coated vesicles, and becomes sequestered in intracellular vacuoles and lysosomes. Considerable biochemical evidence indicates that the low lysosomal pH triggers the fusion of the viral membrane with the intracellular vacuolar membrane leading to the release of the nucleocapsid into the cytoplasm. A similar pathway for the entry of macromolecules into lysosomes utilizing receptor-mediated endocytosis has been found for low density lipoprotein and several other polypeptides (reviewed by Goldstein et al., 1979).

Once inside the cytoplasm the genomic 49S RNA is somehow uncoated (presumably involving the dissociation of capsid protein) and serves as the mRNA for the viral RNA replicase. Naked 49S RNA is also infectious, but the efficiency is greatly reduced when compared with intact virions. The (+)-stranded 49S RNA then serves as a template for the production of complementary (-)-stranded 49S RNA (Strauss and Strauss, 1972b). This RNA is transcribed to produce additional genomic 49S RNA as well as a subgenomic 26S RNA consisting of the 3' terminal one-third of the 49S RNA (Simmons and Strauss, 1972a,b; Ou et al., 1981). The 26S RNA serves as the mRNA for the virion structural polypeptides (Simmons and Strauss, 1974).

The 26S RNA is translated from a single initiation site (Cancedda et al., 1975) on membrane-bound polyribosomes (Wirth et al., 1977), and the nascent polyprotein

is processed by proteolytic cleavage (Strauss and Strauss, 1977; see Chapter 3). The genes are translated in the order 5'-C-PE2-E1-3' (PE2 is the precursor to E2) (Garoff et al., 1980a,b; Schlesinger and Kääriäinen, 1980; see Chapter 3). The capsid protein is cleaved while nascent and rapidly associates with genomic 49S RNA to form cytoplasmic nucleocapsids (Söderlund, 1973). Several lines of evidence suggest that the proteolytic activity responsible for this cleavage may reside in the capsid protein itself (Simmons and Strauss, 1974; Scupham et al., 1977; Aliperti and Schlesinger, 1978).

Both of the virion glycoprotein precursors, PE2 and E1, are cotranslationally translocated across the membrane of the rough endoplasmic reticulum (Garoff et al., 1978; Bonatti et al., 1979) by a mechanism which has been the subject of many recent investigations. Secreted and bitopic integral membrane proteins often contain a short (15-20 amino acids) NH_2-terminal hydrophobic peptide, or "signal sequence", which interacts with protein components in the membrane of the rough endoplasmic reticulum and initiates their cotranslational translocation across the membrane (reviewed by Blobel et al., 1979; see also Davis and Tai, 1980; Wickner, 1980; Inouye and Halegoua, 1980; Emr et al., 1980). In most cases, the signal sequence is cleaved by an enzyme(s) called signalase located on the cisternal side of the rough endoplasmic reticulum (Blobel et al., 1979) before synthesis of the polypeptide is complete, and rapidly degraded (Habener et al., 1979). The cleavage of the capsid protein enables the nascent NH_2 terminus of PE2 to function as a signal sequence. Mutants defective in this cleavage fail to insert PE2 into the lumen of the rough endoplasmic reticulum and accumulate a polyprotein precursor which can therefore be degraded by protease treatment of infected cell microsomes (Wirth et al., 1979). The NH_2-terminal hydrophobic segment of PE2 differs from most other known signal sequences, in that it is not cleaved (Bonatti and Blobel, 1979; Bell et al., 1981). PE2 is cleaved 10-20 minutes after its synthesis has been completed but this cleavage occurs during the

final stages of viral maturation and apparently involves a protease with a different specificity than that of signalase (see below). In addition, there is an asparagine-linked glycosylation site in the middle of the putative hydrophobic signal sequence (Bell et al., 1981; J. Mayne, personal communication; see Chapter 3).

Immediately 3' to the coding sequence for PE2, the 26S RNA encodes a hydrophobic peptide, 55 amino acids in length (see Chapter 3), which has been recently isolated and characterized called the 6K protein (Welch and Sefton, 1979, 1980; Welch et al., 1981). This hydrophobic peptide presumably functions as the signal sequence for the insertion of E1. The presence of a separate signal sequence for E1, distinct from that for PE2, is supported by the identification of a ts mutant of Semliki Forest virus which fails to cleave the capsid protein from PE2 at the restrictive temperature but allows the normal insertion of E1 into the lumen of the rough endoplasmic reticulum (Hashimoto et al., 1981).

The timing of the proteolytic cleavages and location of the protease(s) responsible for them are of particular importance since they should give important insights into the topography of the glycoprotein precursors during and after their synthesis. PE2 has been shown to span the bilayer, and contains approximately 30 amino acids COOH-terminal to its transmembrane segment (Wirth et al., 1977; Garoff et al., 1978; Bonatti et al., 1979; Chapter 4). Based on this structure, models for the insertion and cleavage of the glycoprotein precursors can be proposed. While the identity of the protease(s) responsible for the two cleavages separating PE2 and E1 is unknown, both cleavages occur after alanine (see Chapter 3) and signalase appears to cleave after amino acids with short side chains (Blobel et al., 1979). If signalase is responsible for both cleavages, and is localized exclusively on the cisternal side of the rough endoplasmic reticulum (Blobel et al., 1979), then PE2 must at least transiently span the bilayer twice in order for the PE2-6K cleavage to occur in the lumen of the rough endoplasmic reticulum (see Chapter 3). Alternatively, if a proteolytic enzyme

localized on the cytoplasmic side of the bilayer is responsible for the PE2-6K cleavage, then PE2 would be required to cross the membrane only once. If the PE2-6K cleavage does occur on the cytoplasmic side, then the newly exposed NH_2-terminus of the 6K protein could serve as a signal sequence for the insertion of E1. However, kinetic data on the appearance of the glycoproteins suggest that the cleavage separating PE2 and the 6K protein is delayed and occurs after a significant portion of E1 has been synthesized and transferred into the lumen of the rough endoplasmic reticulum (Welch et al., 1981). These data imply that the 6K protein may function as a signal sequence for the insertion of E1 before it is cleaved from PE2. This would be functionally analogous to the secreted chick oviduct protein, ovalbumin, which contains an internal signal peptide (Lingappa et al., 1979), but in contrast to the 6K protein, this peptide is not cleaved from the mature protein. Future experiments will be required to distinguish between these various alternative pathways for the cotranslational membrane insertion and cleavage of the glycoprotein precursors in the rough endoplasmic reticulum.

During or shortly after protein synthesis, mannose-rich oligosaccharide units are added to the glycoprotein precursors (Sefton, 1977). These oligosaccharide units are linked to the polypeptide chain by an N-glycosidic band between asparagine and N-acetyl glucosamine (Burke and Keegstra, 1979; see Chapter 3). The glycoproteins then appear to follow the vectorial route of other secreted and plasma membrane proteins to the Golgi and eventually to the plasma membrane (Palade, 1975; Erwin and Brown, 1980). Recently, clathrin-coated vesicles have been implicated in the transport of the vesicular stomatitis virus glycoprotein from the rough endoplasmic reticulum to the Golgi and from the Golgi to the plasma membrane (Rothman and Fine, 1980). This pathway may be involved in the transport of other membrane glycoproteins, including Sindbis virus glycoprotein precursors. During this time, oligosaccharide chains are trimmed and complex chains are modified by the addition of galactose, fucose and sialic acid. In addition, both Sindbis virus glycoproteins contain covalently

attached fatty acids (Schmidt et al., 1979), which are added post-translationally about 10-20 minutes after the completion of protein synthesis, presumably in the Golgi (Schmidt and Schlesinger, 1980). A similar type of fatty acid binding has also been found for the vesicular stomatitis virus glycoprotein as well as some host cell polypeptides (Schmidt and Schelsinger, 1979), and may be a general characteristic of membrane glycoproteins.

The Sindbis glycoproteins aririve at the plasma membrane and can be incorporated into mature virions about 20 minutes after their synthesis. The cleavage of PE2 to E3 and E2, which also occurs at this time is necessary for the final steps in virus maturation (Strauss and Strauss, 1977). In the case of Sindbis virus, E3 is not found in the mature virion and can be recovered in the culture fluid (Welch and Sefton, 1979; J. Mayne, unpublished). This cleavage appears to be analogous to the processing of such proteins as pro-albumin and pro-parathyroid hormone which are cleaved after double basic amino acids in the Golgi shortly before secretion of the mature protein (reviewed by Dean and Judah, 1980; see Chapter 3). The final events during maturation involve the specific interaction of the nucleocapsid with the cytoplasmic portion of the transmembrane glycoprotein(s) leading to virus budding through the host cell plasma membrane. This interaction is highly specific since host cell glycoproteins are rigorously excluded from mature virions (Strauss, 1978). Furthermore, the budded virion contains equimolar quantities of each structural protein (Schlesinger et al., 1972; Garoff et al., 1974; see Chapter 4).

The subsequent chapters of this thesis deal with the determination of the primary structure and topography of the structural proteins of Sindbis virus. In Chapter 2, a technique is presented which was developed for sequence analysis of purified single-stranded RNAs, and used to obtain nearly the entire sequence of Sindbis 26S RNA. Together with protein sequence data, this has led to the precise localization of the structural protein genes, the determination of the complete polyprotein primary

sequence, and the definition of the cleavage sites involved during translation and maturation of the structural proteins (Chapter 3). Chapter 4 presents a study of the orientation of the virion glycoproteins with respect to the lipid bilayer. These results showed that both glycoproteins are anchored in the bilayer by COOH-terminal sequences of hydrophobic amino acids, but that if E1 spans the bilayer, at most a few amino acids are exposed on the cytoplasmic face of the membrane. In contrast, E2 contains an additional 33 amino acids COOH-terminal to its transmembrane anchor some of which may be capable of forming a specific interaction with the nucleocapsid during budding. Since the structure of E1 suggests that it may not directly interact with the nucleocapsid, a stable E1-E2 complex in the plasma membrane would assure a mature virion with equimolar quantities of each glycoprotein. Chapter 5 describes the production of antisera specific for each structural protein which was then used to study the quaternary structure of the virion glycoproteins and their precursors. It was found that E1 and E2 are associated in the infected cell as well as in the mature virion, and that this complex is stable even in the absence of the nucleocapsid.

References

Aliperti, G., and Schlesinger, M. J. (1978) Evidence for an autoprotease activity of Sindbis virus capsid protein. Virology **90**, 366-369.

Bell, J. R., Rice, C. M., Hunkapiller, M. W., and Strauss, J. H. (1981) The N-terminus of PE2 in Sindbis virus infected cells. Submitted for publication.

Birdwell, C. R. and Strauss, J. H. (1974) Distribution of receptor sites for Sindbis virus on the surface of chicken and BHK cells. J. Virol. **14**, 672-678.

Blobel, G., Walter, P., Chang, C., Goldman, B. M., Erickson, A. H., and Lingappa, V. R. (1979) Translocation of proteins across membranes: the signal hypothesis and beyond. In "Symposia of the Society for Experimental Biology," No. 33, pp. 9-36.

Bonatti, S. and Blobel, G. (1979) Absence of a cleavable signal sequence in Sindbis virus glycoprotein E2. J. Biol. Chem. **254**, 12261-12264.

Bonatti, S., Cancedda, R., and Blobel, G. (1979) Membrane biogenesis: in vitro cleavage core glycosylation, and integration into microsomal membranes of Sindbis virus glycoproteins. J. Cell Biol. **80**, 219-224.

Burke, D. and Keegstra, K. (1979) Carbohydrate structure of Sindbis virus glycoprotein E2 from virus grown in hamster and chicken cells. J. Virol. **29**, 546-554.

Cancedda, R., Villa-Komaroff, L., Lodish, H. F., and Schlesinger, M. (1975) Initiation sites for translation of Sindbis virus 42S and 26S messenger RNAs. Cell **6**, 215-222.

Compans, R. W. (1971) Location of the glycoproteins in the membrane of Sindbis virus. Nature New Biol. **229**, 114-116.

Davis, B. D. and Tai, P.-C. (1980) The mechanism of protein secretion across membranes. Nature (London) **283**, 433-438.

Dean, R. T. and Judah, J. D. (1980) Post-translational proteolytic processing of polypeptides. In "Comprehensive Biochemistry," eds. Florkin, M. and Neuberger, A. (Elsevier, Amsterdam), Vol. 19B, pt. 1, pp. 233-298.

Emr, S. D., Hall, M. N., and Silhavy, T. J. (1980) A mechanism of protein localization: the signal hypothesis and bacteria. J. Cell Biol. **86**, 701-711.

Erwin, C. and Brown, D. T. (1980) Intracellular distribution of Sindbis virus membrane proteins in BHK-21 cells infected with wild-type virus and maturation-defective mutants. J. Virol. **36**, 775-786.

Fan, D. P. and Sefton, B. M. (1978) The entry into host cells of Sindbis virus, vesicular stomatitis virus, and Sendai virus. Cell **15**, 985-992.

Garoff, H., Simons, K., and Renkonen, O. (1974) Isolation and characterization of the membrane proteins of Semliki Forest virus. Virology **61**, 493-504.

Garoff, H., Simons, K., and Dobberstein, B. (1978) Assembly of the Semliki Forest virus membrane glycoproteins in the membrane of the endoplasmic reticulum in vitro. J. Mol. Biol. **124**, 587-600.

Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H., and Delius, H. (1980a) Nucleotide sequence of the cDNA coding for the Semliki Forest virus membrane glycoproteins. Nature (London) **288**, 236-241.

Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H., and Delius, H. (1980b) The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino terminal region. Proc. Natl. Acad. Sci. USA **77**, 6376-6380.

Goldstein, J. L., Anderson, R. G. W., and Brown, M. S. (1979) Coated pits, coated vesicles, and receptor mediated endocytosis. Nature (London) **279**, 679-685.

Habener, J. F., Rosenblatt, M., Dee, P. C., and Pott, J. T. (1979) Cellular processing of pre-proparathyroid hormone invovles rapid hydrolysis of the leader sequence. J. Biol. Chem. **254**, 10596-10599.

Hashimoto, K., Erdei, S., Kääriäinen, L., Keränen, S., and Saraste, J. (1981) Evidence for a separate signal sequence for the carboxyterminal envelope glycoprotein E1 of Semliki Forest virus. J. Virol., submitted.

Helenius, A., Morein, B., Fries, E., Simons, K., Robinson, P., Schirrmacher, V., Terhorst, C., and Strominger, J. L. (1978) Human (HLA-A and HLA-B) and murine (H-2K and H-2D) histocompatibility antigens are cell surface receptors for Semliki Forest virus. Proc. Natl. Acad. Sci. USA **75**, 3846-3850.

Helenius, A., Kartenbeck, J., Simons, K., and Fries, E. (1980) On the entry of Semliki Forest virus into BHK-21 cells. J. Cell Biol. **84**, 404-420.

Inouye, M. and Halegoua, S. (1980) Secretion and membrane localization of proteins in Escherichia coli. CRC Crit. Rev. Biochem. **7**, 339-371.

Kääriäinen, L. and Söderlund, H. (1978) Structure and replication of alphaviruses. Curr. Top. Microbiol. Immunol. **82**, 15-69.

Lingappa, V. R., Lingappa, J. R., and Blobel, G. (1979) Chicken ovalbumin contains an internal signal sequence. Nature (London) **281**, 117-121.

Murphy, F. A. (1980) "Togavirus morphology and morphogenesis" In The Togaviruses, Schlesinger, R. W., ed. (Academic Press, New York) pp. 241-316.

Ou, J.-H., Strauss, E. G., and Strauss, J. H. (1981) Comparative studies of the 3'-terminal sequences of several alphavirus RNAs. Virology **109**, 281-289.

Palade, G. (1975) Intracellular aspects of the process of protein synthesis. Science **189**, 347-358.

Rothman, J. E. and Fine, R. E. (1980) Coated vesicles transport newly synthesized membrane glycoproteins from the endoplasmic reticulum to plasma membrane in two successive stages. Proc. Natl. Acad. Sci. USA **77**, 780-784.

Schlesinger, M. J., Schlesinger, S., and Burge, B. W. (1972) Identification of a second glycoprotein in Sindbis virus. Virology **47**, 539-541.

Schlesinger, M. J. and Kääriänen, L. (1980) "Translation and processing of alphavirus proteins" in The Togaviruses, Schlesinger, R. W., ed. (Academic Press, New York) pp. 371-392.

Schmidt, M. F. G. and Schlesinger, M. J. (1979) Fatty acid binding to vesicular stomatitis virus glycoprotein: a new type of post-translational modification of the viral glycoprotein. Cell **17**, 813-819.

Schmidt, M. F. G., Bracha, M., and Schlesinger, M. J. (1979) Evidence for the covalent attachment of fatty acids to Sindbis virus glycoproteins. Proc. Natl. Acad. Sci. USA **76**, 1687-1691.

Schmidt, M. F. G. and Schlesinger, M. J. (1980) Relation of fatty acid attachment to the translation and maturation of vesicular stomatitus virus and Sindbis virus membrane glycoproteins. J. Biol. Chem. **255**, 3334-3339.

Scupham, R. K., Jones, K. J., Sagik, B. P., and Bose, H. R., Jr. (1977) Virus-directed post-translational cleavage in Sindbis virus-infected cells. J. Virol. **22**, 568-571.

Sefton, B. M. (1977) Immediate glycosylation of Sindbis virus membrane proteins. Cell **10**, 659-668.

Simmons, D. T. and Strauss, J. H. (1972a) Replication of Sindbis virus. I. Relative size and genetic content of 26S and 49S RNA. J. Mol. Biol. **71**, 599-613.

Simmons, D. T. and Strauss, J. H. (1972b) Replication of Sindbis virus. II. Multiple forms of double-stranded RNA isolated from infected cells. J. Mol. Biol. **71**, 615-631.

Simmons, D. T. and Strauss, J. H. (1974) Translation of Sindbis virus 26S RNA and 49S RNA in lysates of rabbit reticulocytes. J. Mol. Biol. **86**, 397-409.

Simons, K., Garoff, H., Helenius, A., and Ziemiecki, A. (1978) "The structure and assembly of the membrane of Semliki Forest virus." In Frontiers of Physio-chemical Biology, B. Pullman, ed. (Academic Press, New York) pp. 397-407.

Söderlund, H. (1973) Kinetics of formation of Semliki Forest virus nucleocapsid. Intervirology **1**, 354-361.

Strauss, J. H. and Strauss, E. G. (1977) "Togaviruses" In Molecular Biology of Animal Viruses, Nayak, D., ed. Vol. 1, pp. 111-166.

Strauss, E. G. (1978) Mutants of Sindbis virus. III. Host polypeptides present in purified HR and ts 103 virus particles. J. Virol. **28**, 466-474.

Welch, W. J. and Sefton, B. M. (1979) Two small virus-specific polypeptides are produced during infection with Sindbis virus. J. Virol. **29**, 1186-1195.

Welch, W. J. and Sefton, B. M. (1980) Characterization of a small nonstructural viral polypeptide present late during infection of BHK cells by Semliki Forest virus. J. Virol. **33**, 230-237.

Welch, W. J., Sefton, B. M., and Esch, F. S. (1981) Amino terminal sequence analysis of alphavirus polypeptides. J. Virol., submitted.

Wickner, W. (1980) Assembly of proteins into membranes. Science **210**, 1186-1195.

Wirth, D. F., Katz, F., Small, B., and Lodish, H. F. (1977) How a single Sindbis virus mRNA directs the synthesis of one soluble and two integral membrane glycoproteins. Cell **10**, 253-263.

Wirth, D., Lodish, H. F., and Robbins, P. W. (1979) Requirements for the insertion of the Sindbis envelope glycoproteins into the endoplasmic reticulum membrane. J. Cell Biol. **81**, 154-162.

CHAPTER 2

Synthesis, cleavage, and sequence analysis of cDNA complementary to the 26S mRNA of Sindbis virus

CHARLES M. RICE and JAMES H. STRAUSS

Journal of Molecular Biology, in press.

ABSTRACT

Conditions for synthesis of long cDNA transcripts of Sindbis virus 26S and 49S RNA in high yield have been developed. This single-stranded cDNA could be cut with Hae III, Hha I, Rsa I, or Taq I to give reproducible patterns of discrete, virus-specific fragments which were suitable for subsequent end-labeling and direct sequence analysis. Using these methods we present the strategy used for obtaining nearly the entire 26S RNA sequence from cDNA synthesized in vitro. This approach should prove useful for sequence analysis of any purified RNA available in microgram quantities.

INTRODUCTION

Sindbis virus, and the closely related Semliki Forest virus (SFV) have been extensively studied in several laboratories as model systems for alphavirus maturation and glycoprotein biosynthesis. These simple viruses contain a lipid envelope with two or three spike glycoproteins (E1, E2, and in the case of SFV, E3), which surrounds an icosahedral nucleocapsid consisting of the capsid protein (C) and the 49S genomic RNA. They mature by budding through the host cell plasma membrane (Strauss and Strauss, 1977), and contain exclusively viral encoded glycoproteins (Strauss, 1978). Thus, the interaction between the transmembrane viral glycoproteins and the cytoplasmic nucleocapsid is highly specific in these viruses. At the onset of virus infection in vertebrate cells, host cell macromolecular synthesis is shut off (Wengler, 1980), and large quantities of two virus-specific mRNAs are produced. The genomic 49S RNA serves as the mRNA for the nonstructural polypeptides. A 3' terminal subgenomic 26S RNA is translated from a single initiation site (Cancedda et al., 1975) to produce the virion structural polypeptides in the order 5'-C-E3-E2-E1-3' (Schlesinger and Kääriäinen, 1980). Primary amino acid sequence data for the NH_2-terminal regions of the Sindbis structural proteins and of some precursors have recently been obtained (Bell et al., 1978; Bell et al., in preparation; Rice et al., in preparation). However, NH_2-terminal blockage of the capsid protein (Boege et al., 1980; Bell and Strauss, in preparation) and the scarcity of intermediate cleavage products have made the complete polyprotein sequence difficult to determine. For these reasons and others, it was of interest to determine the 26S RNA sequence. In conjunction with available protein sequence data, the RNA sequence would allow the precise localization of the structural protein genes, the determination of the complete polyprotein primary sequence, and the definition of the cleavage sites involved during translation and maturation of the structural proteins. In addition, examination and

comparison of related alphavirus genomes for common sequences or secondary structures should give useful insights into their possible role in RNA replication, translational regulation, and encapsidation. As an alternative method for rapid sequence determination, we have sequenced single-stranded cDNA restriction fragments complementary to Sindbis 26S RNA (26S cDNA). This approach is similar to the one used for the sequence determination of the 5'-untranslated regions of human α- and β-globin mRNAs (Chang et al., 1977).

This method requires only a few micrograms of purified RNA, and because molecular cloning is not involved, there is no chance of selecting a minor variant in the population. In addition, this approach should be extremely valuable for the rapid determination of sequence changes in Sindbis ts mutants of which a large, well characterized catalogue exists (Strauss and Strauss, 1980) or for sequence comparisons of closely related virus strains. In this paper we present the methods used for cleavage of single-stranded cDNA into discrete fragments which were then end-labeled, isolated, and sequenced using the basic-specific chemical cleavage method of Maxam and Gilbert (1980). From the overlapped sequence we have verified that Hae III, Hha I, Rsa I, and Taq I have the same specificity on single-stranded DNA as doublestranded DNA, but that additional sites appear to be cleaved.

MATERIALS AND METHODS

Virus growth and RNA purification. The HR (large plaque) strain of Sindbis virus (Burge and Pfefferkorn, 1966) was grown in monolayers of primary chicken embryo fibroblasts as previously described (Pierce et al., 1974). Intracellular 26S RNA was prepared by the method of Ou et al. (1981). Briefly, poly(A) containing RNA was selected from whole cytoplasmic RNA by two passages over oligo(dT) cellulose, and the 26S fraction was pooled after velocity sedimentation on sucrose gradients. The final yield of 26S RNA was between 5-20 μg per 800 cm^2 roller bottle.

Virion 49S RNA was extracted from purified virus (Bell et al., 1979) after denaturation with 0.5% SDS using the phenol/chloroform method of Hsu et al. (1973). After two successive ethanol precipitations RNA was resuspended in distilled water at a concentration of 0.5-1 mg/ml and stored at -70°C.

Molecular weight markers were prepared by digestion of M13 (strain mp 73, from J. Messing) virion DNA (kindly provided by H. V. Huang) with Hae III and 5' end-labeled using T4 polynucleotide kinase (P.L. Biochemicals) and [γ-^{32}P]-ATP (ICN), as described below.

cDNA synthesis. cDNA was synthesized essentially under the conditions of Myers and Spiegelman (1978) with several modifications. Typical reactions contained 50 mM Tris-Cl, pH 8.3, 8 mM $MgCl_2$, 1 mM dithiothreitol, 50 mM KCl, 1 mM of all four deoxynucleotide triphosphates (dNTP), 10 µg/ml oligo(dT)$_{12-18}$ (P.L. Biochemicals) or 300 µg/ml short calf thymus DNA (6-8 nucleotides, from J. Casey), 15-50 µg/ml Sindbis 26S RNA, 4 mM sodium pyrophosphate, and 12 units of avian myeloblastosis virus (AMV) reverse transcriptase (kindly provided by the Office of Program Resources and Logistics, Viral Cancer Program, Viral Oncology, Division of Cancer Cause and Prevention, National Cancer Institute) per microgram of template RNA. cDNA for subsequent sequence analysis was labeled with trace amounts of ^{3}H-TTP (Amersham Searle). For preparation of uniformly labeled cDNA for restriction analysis 50 µCi of each [α-^{32}P]-dNTP (400 Ci/mmole) (Amersham Searle) was included and the unlabeled dNTP concentration lowered to 0.2 mM. Occasionally, 4 mM ribonucleoside-vanadyl complexes (from Kai Zinn) and additional 4 mM $MgCl_2$ were included in the reaction mixture (see Results). The template RNA was heated to 56°C for 2 minutes and quick chilled in ice water before addition to the mixture. Synthesis was allowed to proceed for 20-60 minutes at 42.5°C and stopped by the addition of Na_2 EDTA to 25 mM. Incorporation was measured by spotting duplicate samples of the reaction mixture onto Whatman 3MM discs followed by precipitation with

trichloracetic acid and liquid scintillation counting. Products were sized on vertical alkaline agarose gels as described by McDonell et al. (1977).

The reverse transcription mixture was extracted once with phenol and twice with chloroform:isoamyl alcohol (100:1). 25 µg carrier tRNA (P.L. Biochemicals) and sodium acetate, pH 5.8, to 0.2 M were added and the mixture was precipitated with 2.5 volumes of cold 100% ethanol. After pelleting and resuspension in distilled water, the cDNA-RNA hybrids were used directly for restriction enzyme analysis or alternatively the RNA strand was hydrolyzed in 0.1 M NaOH for 60 minutes at 60°C. After hydrolysis samples for subsequent end-labeling and sequence analysis were chilled and diluted with 3 volumes of distilled water, neutralized with HCl, and passed over a Biogel A5M (BioRad) column equilibrated in 25 mM Tris-Cl, pH 7.4, 1 mM Na_2 EDTA, 0.02% sodium dodecyl sulfate (SDS) to remove small oligonucleotides. The excluded peak fractions were pooled, lyophilized, and resuspended in 100 µl distilled water. After addition of sodium acetate to 0.2 M, and magnesium acetate to 20 mM, the cDNA was precipitated with 2.5 volumes of cold 100% ethanol. This precipitation was repeated to assure complete removal of the SDS. cDNA samples were resuspended in distilled water and stored at -20°C.

Restriction enzyme digestion and end labeling. Restriction endonucleases Hae III, Hha I, Rsa I, and Taq I were all purchased from New England Biolabs. Reaction conditions were essentially as recommended by the manufacturer. An excess 10 to 50 fold of the restriction enzymes was required for nearly complete digestion of single-stranded cDNA or cDNA-RNA hybrids. The amount of enzyme used was empirically determined for each batch because of the large variability in the levels of nonspecific nuclease activity in different batches of these and other restriction endonucleases from various commercial sources (see Results). Taq I was preincubated for 15 minutes at 70°C to inactivate nonspecific nucleases (Sato et al., 1977) before

addition of the substrate. Taq I digestions were done at 56°C or 65°C, all others were incubated at 37°C. After 1 hour the reaction was stopped by the addition of Na_2 EDTA to 25 mM.

Samples for 5' or 3' end-labeling were extracted with phenol, then with chloroform, and ethanol precipitated as described above (omitting the carrier tRNA). Alternatively, restriction digests in low-salt buffers were sometimes used directly after dilution and adjustment to pH 8.3 with Tris-Cl. In either case, 37 units of bacterial alkaline phosphatase (Bethesda Research Labs) were added to 0.5-1 µg of cDNA in a final volume of 40-60 µl, and the mixture was incubated at 65°C for 30 minutes. After the addition of Na_2 EDTA to 25 mM the mixture was extracted with phenol and chloroform followed by ethanol precipitation as described above. 5' end-labeling was done essentially as described by Maxam and Gilbert (1980) with 7.5 units of T4 polynucleotide kinase (P.L. Biochemicals) per microgram of cDNA and 1 mCi of [γ-^{32}P]-ATP (ICN, >7000 Ci/mmole) in a final reaction volume of 20 µl. 3' end-labeling with terminal deoxynucleotidyl transferase (ribosubstitution grade, Bethesda Research Labs) and [α-^{32}P]-ATP was done as previously described (Maxam and Gilbert, 1980). Base treatment after 3' end-labeling was either omitted or done as described above for removal of the RNA from cDNA-RNA hybrids.

Oligo(dA) cellulose chromatography. To obtain the sequence of cDNA complementary to the 3' end of the 26S RNA adjacent to the poly(A) tail oligo(dA) cellulose (Collaborative Research) chromatography was used for an additional purification step. Oligo(dT)$_{12-18}$ primed cDNA digested with Rsa I or Hha I and 3' end-labeled with terminal deoxynucleotidyl transferase was passed over oligo(dA) cellulose (see the procedure for oligo(dT) cellulose chromatography in Ou et al., 1981). The bound fraction was ethanol precipitated and purified on a preparative gel as described below. The column retained a population of fragments with a unique 3' end (if restriction

enzyme cleavage was complete) and a heterogeneous 5' end (due to random oligo(dT) priming along the poly(A) tail during cDNA synthesis).

Separation of cDNA restriction fragments. Analytical and preparative gels were 6% acrylamide (acrylamide:bisacrylamide, 20:1), 8.3 M urea, 2 mM Na_2EDTA, and 100 mM Tris-borate, pH 8.3, prepared and run essentially as described by Maxam and Gilbert (1980). Samples in 80% deionized formamide, 50 mM Tris-borate, pH 8.3, and 1 mM Na_2EDTA, were denatured at 90°C for 2 minutes and quick chilled prior to electrophoresis. Gel dimensions were 30 cm x 40 cm x 0.5 mm or 30 cm x 80 cm x 0.5 mm. Only the notched plate was siliconized. After electrophoresis analytical gels were transferred to Whatman 3MM filter paper, covered with a thin plastic film (910, Reynolds) and exposed at -70°C using prefogged Kodak X-Omat R film and a Cronex Lightning Plus (DuPont) intensifying screen (Laskey and Mills, 1977), or at -20°C with unfogged film and no intensifying screen. Preparative gels were covered and exposed at room temperature. Fragments for sequencing were excised and eluted for 24-72 hours into 1 ml of 0.6 M sodium acetate, 0.1 M Tris-Cl, pH 8.0, 2 mM Na_2EDTA, 25 µg/ml yeast tRNA, or by electroelution overnight. The DNA was separated from the gel slice by centrifugation through a siliconized glass wool plug, and precipitated with three volumes of cold 100% ethanol. The pellets were collected by centrifugation, redissolved in 200 µl 0.1 M ammonium acetate and the DNA precipitated a second time with 2.5 volumes of 100% ethanol. After resuspension in distilled water, each sample was dispensed into four Eppendorf tubes, frozen, and lyophilized before sequencing. Using these preparative methods we could prepare up to 20 or 30 end-labeled fragments for sequencing from a single preparation of restriction endonuclease cleaved cDNA (about 0.5 µg cDNA derived from 1-2 µg 26S RNA).

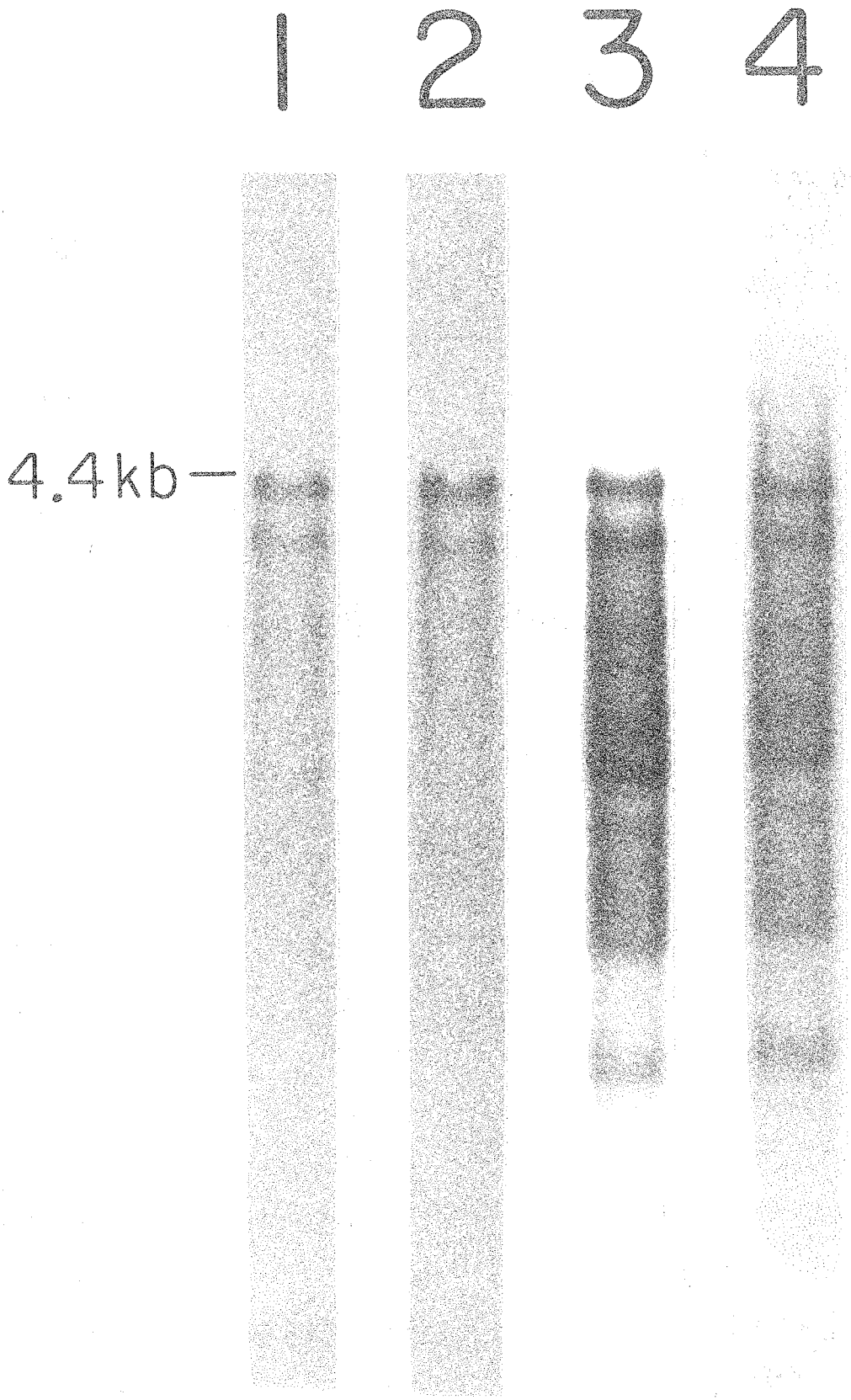
Nucleotide sequence determination. Base-specific chemical cleavage of the

end-labeled fragments was done essentially as described by Maxam and Gilbert (1980) or as modified by Smith and Calvo (1980). Four reactions were used: either C, C + T, A > C, and G > A or C, C + T, G + A, and G. For reading 300 nucleotides or less, the G + A reaction (final volume 23 µl) contained 3 µl of 1 M pyridinium formate, pH 2.0, and was incubated at 37°C for 10 minutes. For longer sequences 2 µl pyridinium formate and a 5 minute incubation at 37°C was used. Sequencing gels (Maxam and Gilbert, 1980) were either 5%, 8%, 20%, or 25% acrylamide and run at high temperatures ($\geq$50°C). Samples in 80% deionized formamide were denatured at 90°C for 2 minutes and quick chilled prior to electrophoresis. After electrophoresis, the gels were transferred to either Whatmann 3MM filter paper or blotting paper and exposed to X-ray film as described above.

RESULTS

Synthesis of cDNA. Optimal conditions for the synthesis of long cDNA to Sindbis virus 26S and 49S RNAs were found to be similar to those used for poliovirus (Kacian and Myers, 1976; Myers and Spiegelman, 1978), and several other mRNAs (Zain et al., 1979; Devos et al., 1979; Buell et al., 1978). Increasing the reaction temperature from 37°C to 42.5°C resulted in a larger proportion of full length oligo(dT) primed cDNA to 26S RNA (data not shown). Preliminary experiments (data not shown) showed that the AMV reverse transcriptase was contaminated with ribonuclease, and that increasing the amount of enzyme in the reaction eventually lead to a decrease in the average length of the cDNA synthesized using an oligo(dT) primer. 12 units of AMV reverse transcriptase per µg of 26S RNA was found to be optimal for production of full length cDNA. In addition, the inclusion of 4 mM ribonucleoside-vanadyl complexes (Berger and Birkenmeier, 1979) and an additional 4 mM Mg^{++} increased the yield of full length cDNA approximately two-fold (Fig. 1). Using these conditions from 30 to 60% of the input RNA was converted to cDNA.

Figure 1. Alkaline agarose gel patterns of oligo(dT) primed cDNA made to Sindbis virus 26S (lanes 1-3) or 49S (lane 4) RNA. Conditions for the cDNA reactions were as described in Materials and Methods except that lane 2 contained 4 mM ribonucleoside-vanadyl complexes, and lane 3 contained only 4 units of AMV reverse transcriptase per microgram of 26S RNA. The marker at 4.4 kilobases was the position of end-labeled linearized pBR322.


1
2
3
4
4.4kb

Since cleavage of a cDNA into discrete fragments for sequencing depends on the presence of at least two cleavage sites, it was important to produce long cDNA transcripts. For short poly(A) containing RNAs this can usually be accomplished using an oligo(dT) primer. However, for longer RNAs inefficiently copied from the 3' end or lacking a poly(A) tail other priming sites must be used. In order to obtain a uniform distribution of 26S sequences, cDNA for sequencing was usually randomly primed using short calf thymus DNA. Thus, it was important to adjust the primer to template ratio such that long cDNA was produced. We found that ~6 μg calf thymus primer/μg template RNA gave predominantly long cDNA (>500 nucleotides, data not shown) suitable for subsequent digestion by frequently cutting restriction endonucleases with four base recognition sequences (see below).

Figure 1 shows a comparison of oligo(dT) primed cDNA made from intracellular 26S RNA and virion 49S RNA. In either case the cDNA formed a discrete banding pattern terminating abruptly at about 4.2 kb (approximately the size of 26S RNA). The cDNA made from 49S virion RNA had a small proportion of longer transcripts. These discrete cDNA species could have resulted from premature termination of cDNA synthesis due to RNA secondary structure (Kacian and Myers, 1976; Owens and Cress, 1980) or sites in the template RNA particularly susceptible to ribonuclease. Since the 26S RNA is a 3' terminal subgenomic fraction of the virion 49S RNA (Simmons and Strauss, 1972; Kennedy, 1976; Ou et al., 1981), the pattern of oligo(dT) primed cDNA products smaller than 4.2 kb from either the 26S or 49S RNA was virtually identical (Fig. 1). The absence of extra bands in the 49S cDNA (smaller than 4.2 kb) indicated that most of the oligo(dT) primed cDNA synthesis began at the 3' poly(A) tail, as expected, rather than at other potential priming sites elsewhere in the genome. We have been unable to obtain a high yield of oligo(dT) primed cDNA transcripts from 49S RNA longer than 4.2 kb. The molecular basis and significance, if any, of this apparent "strong stop" for reverse transcriptase in 49S RNA is at present unknown.

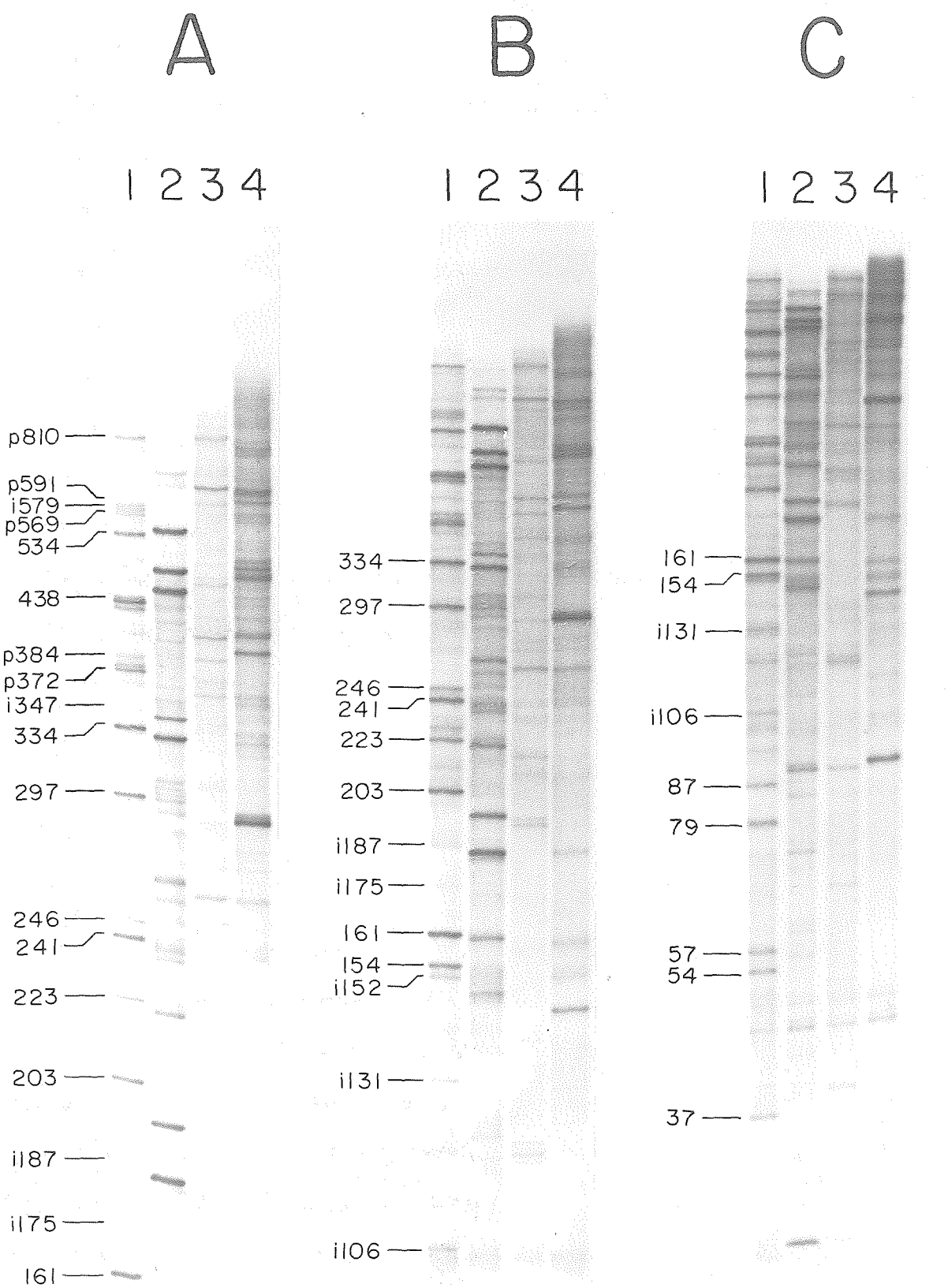
Restriction enzyme cleavage of cDNA. In order to use the in vitro synthesized cDNA for direct sequence analysis, it was necessary to find Type II restriction enzymes which would produce discrete fragments with unique 5' and 3' termini. Hae III, Hha I, Hinf I, Hpa II, and Mbo I have been reported to cleave single-stranded DNA (Horiuchi and Zinder, 1975; Blakesley and Wells, 1975; Godson and Roberts, 1976). Additional enzymes, including Taq I, have been found which cleave at least the DNA strand of cDNA-RNA hybrids (Molloy and Symons, 1980). We have found that Rsa I and Taq I also cleave single-stranded DNA into discrete fragments.

In Fig. 2 the restriction patterns obtained by digestion of cDNA from 26S RNA by Hae III, Hha I, Rsa I, and Taq I are shown. All four of these enzymes also cut the DNA strand of cDNA-RNA hybrids with virtually the same efficiency and specificity found for singlestranded DNA (data not shown). In addition, the fragments produced by these enzymes were suitable for subsequent end-labeling and sequence analysis (see below).

Extensive overdigestion of the cDNA, especially with Hha I, resulted in a low yield of discrete fragments due to nonspecific endonuclease and exonuclease activity. Hae III, Hha I, Rsa I, and Taq I were used in 40, 20, 50, and 10 fold excess, respectively, over the amount of enzyme required to cut an equivalent amount of duplex lambda DNA to completion. It should be emphasized, however, that these digestion conditions were empirically optimized for each batch of enzyme (data not shown) and are not generally applicable.

Since the 26S RNA template is about 4.2-4.5 kb (Strauss and Strauss, 1977; C. M. Rice, unpublished data), we expected a complete restriction digest to yield a set of fragments whose sizes summed to about 4.2 kb, as well as an approximate linear relationship between fragment size and incorporated label (using uniformly labeled cDNA as a substrate). Inspection of the Hae III digest of uniformly labeled cDNA to 26S RNA in Fig. 2 revealed that the fragments were not present in equimolar

Figure 2. Restriction endonuclease digestion of randomly primed cDNA synthesized from Sindbis 26S RNA. Hae III (lane 1), Taq I (lane 2), Hha I (lane 3), and Rsa I (lane 4) fragments were prepared and separated on 40 cm long 6% acrylamide gels as described in Materials and Methods. The xylene cyanol tracking dye was run to 60 cm (panel A), 40 cm (panel B), or 20 cm (panel C). The sizes of sequenced Hae III fragments (see Fig. 5, Table 1) are given in nucleotides.


A
B
C
1 2 3 4
1 2 3 4
1 2 3 4
p810
p591
i579
p569
534
438
p384
p372
i347
334
297
246
241
223
203
i187
i175
161
334
297
246
241
223
203
i187
i175
161
154
i152
i131
i106
161
154
i131
i106
87
79
57
54
37

amounts, and that their lengths summed to more than 9 kb. Based on sequence analysis of more than 40 of the Hae III fragments (see below), it has become clear that 26S RNA sequence heterogeneity (largely 49S-specific contaminants), differential transcription of 26S sequences by AMV reverse transcriptase, and partial Hae III digestion were responsible for the complex pattern seen in Fig. 2.

Determination of specificity of restriction fragments. In Fig. 3 we have compared the Hae III restriction pattern of cDNA from intracellular 26S RNA and virion 49S RNA primed with either oligo(dT)$_{12-18}$ or small oligonucleotides from calf thymus DNA. Virtually all (>90%) of the Hae III cleaved cDNA fragments from 26S RNA appeared to be virus-specific, regardless of the primer used. There were additional fragments present in the Hae III digest of calf thymus primed cDNA from virion 49S RNA, since the 26S RNA accounts for only about one third of the 49S sequences (Simmons and Strauss, 1972). As we began sequencing these Hae III fragments (see below) it became clear that virtually all were virus-specific, as the above experiment had suggested, and that such a comparison for the other restriction enzymes (Hha I, Rsa I, and Taq I) was unnecessary. This comparison should prove useful, however, for identifying virus-specific cDNA restriction fragments when only impure preparations of intracellular RNA are available.

Separation and sequence analysis of cDNA restriction fragments. After restriction enzyme cleavage and 3' or 5' end-labeling the single-stranded cDNA restriction fragments were separated on 6% acrylamide sequencing gels (Maxam and Gilbert, 1980). Even large fragments (>300 nucleotides), differing by only a few nucleotides, were usually resolved and could be cleanly excised from the gel for sequence analysis. Figure 4 shows such a preparative gel. As little as 300 ng of cDNA was sufficient to produce a number of end-labeled fragments suitable for sequencing. The selection on oligo(dA) cellulose of oligo(dT) primed cDNA digested with Hha I and 3' end-labeled

Figure 3. Comparison of Hae III fragments from cDNA to Sindbis 26S RNA (lanes 2 and 4) and virion 49S RNA (lanes 3 and 5). The cDNA was synthesized using either oligo(dT) (lanes 2 and 3) or short fragments of calf thymus DNA (lanes 4 and 5) as primers. The fragments were separated on 40 cm long 6% acrylamide gels in which the xylene cyanol was run for 60 cm (panel A), 40 cm (panel B), or 20 cm (panel C). Lane 1 is a 5' end-labeled Hae III digest of M13 (strain mp73). The lengths of sequenced Hae III fragments of the 26S cDNA (see Fig. 5, Table 1) are given as molecular weight markers.

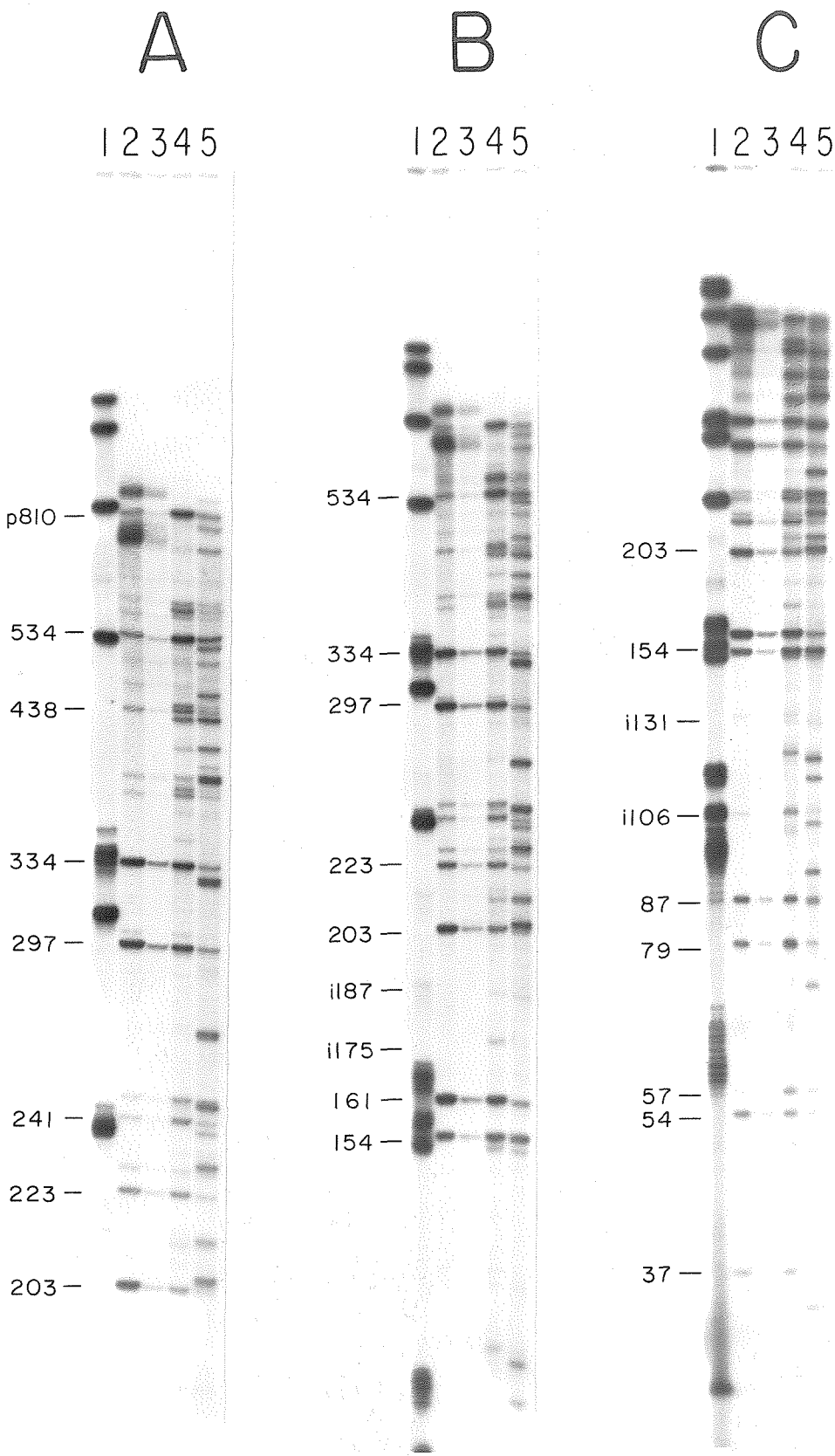
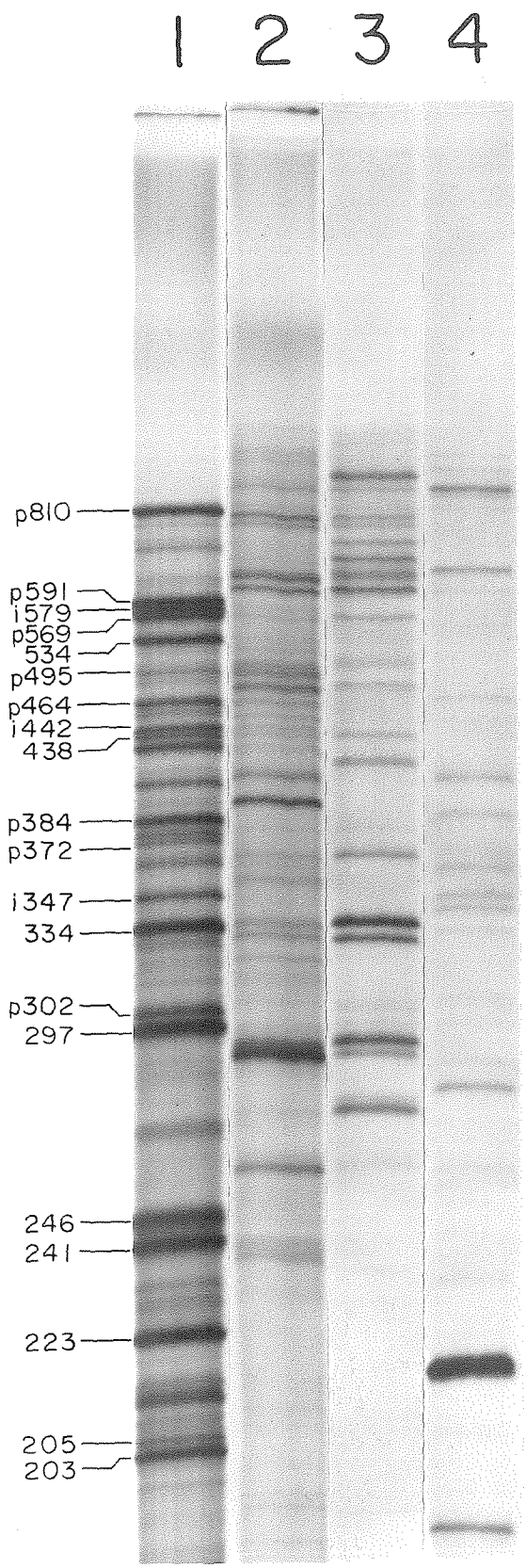

A
1 2 3 4 5
p810
534
438
334
297
241
223
203
B
1 2 3 4 5
534
334
297
223
203
i187
i175
161
154
C
1 2 3 4 5
203
154
i131
i106
87
79
57
54
37

Figure 4. Preparative 40 cm long 6% acrylamide gel of 5' end-labeled (lanes 1 and 2) or 3' end-labeled (lanes 3 and 4) restriction fragments. Randomly primed (lanes 1 and 2) or oligo(dT) primed (lanes 3 and 4) cDNA to Sindbis 26S RNA was digested with Hae III (lane 1), Rsa I (lanes 2 and 3), or Hha I (lane 4) and end-labeled as described in the Materials and Methods. The sizes of Hae III fragments (see Fig. 5, Table 1) are indicated.


1
2
3
4
p810
p591
i579
p569
534
p495
p464
i442
438
p384
p372
i347
334
p302
297
246
241
223
205
203

was useful for obtaining the sequence adjacent to the poly(A) tail. Further purification on sequencing gels of the oligo(dA) bound fragments was necessary since it was difficult to achieve complete digestion of the cDNA with these enzyme preparations without extensive degradation. 5' end-labeling with high specific activity γ-ATP and polynucleotide kinase is very efficient on single-stranded DNA substrates and was especially useful for labeling restriction fragments produced in low yield. More than 80% of the Hae III fragments excised from the preparative gels gave useful sequence data, although several minor bands contained considerable sequence heterogeneity. We are presently developing a two-dimensional separation to improve resolution and therefore increase the proportion of homogeneous fragments for sequencing.

Figure 5 shows a sequencing schematic of the cDNA fragments used in determining the 26S RNA sequence. More than 17,000 nucleotides of sequence data from about 100 different restriction fragments were used to verify and overlap nearly the entire sequence. cDNA sequences were stored and overlapped using a computer program, and the complementary RNA sequence and encoded protein sequences have been presented elsewhere (Rice and Strauss, 1981). The restriction sites for Hae III, Hha I, Rsa I, and Taq I, as well as two areas of secondary structure are shown in Fig. 6 in which the cDNA sequence is numbered from the 5' end (the cDNA sequence begins with the complement of the first residue adjacent to the poly(A) tail). Table 1 summarizes the exact location and size of the restriction fragments which were useful in obtaining the 26S cDNA sequence. Although sequencing gels were run at high temperatures (>50°C) to avoid compression artifacts due to DNA secondary structure (Maxam and Gilbert, 1980), two areas of secondary structure still caused difficulties (see Fig. 6). For example, in the first region of secondary structure which includes the sequence GCGCATACTG (from position 1292 to position 1301 of the cDNA sequence), we are uncertain of the sequence CATA, which could also be read as CAAT. In the latter case, the valine deduced at position 113 of E1 would instead be an aspartic

Figure 5. Sequencing schematic of restriction fragments used to determine the 26S cDNA sequence. The direction and extent of sequence analysis are indicated by arrows and solid lines, respectively. Unsequenced portions are shown by dashed lines. Legitimate and illegitimate restriction endonuclease cleavage sites are indicated by solid and open circles, respectively, at the ends of the fragments. The arrows show the restriction sites for Hae III, Hha I, Rsa I, and Taq I, on duplex DNA. The lengths of complete, partial (denoted by "p"), or illegitimate (denoted by "i") fragments are given in nucleotides. The lengths of illegitimate fragments are approximate and were estimated on gels using sequenced restriction fragments as molecular weight standards.

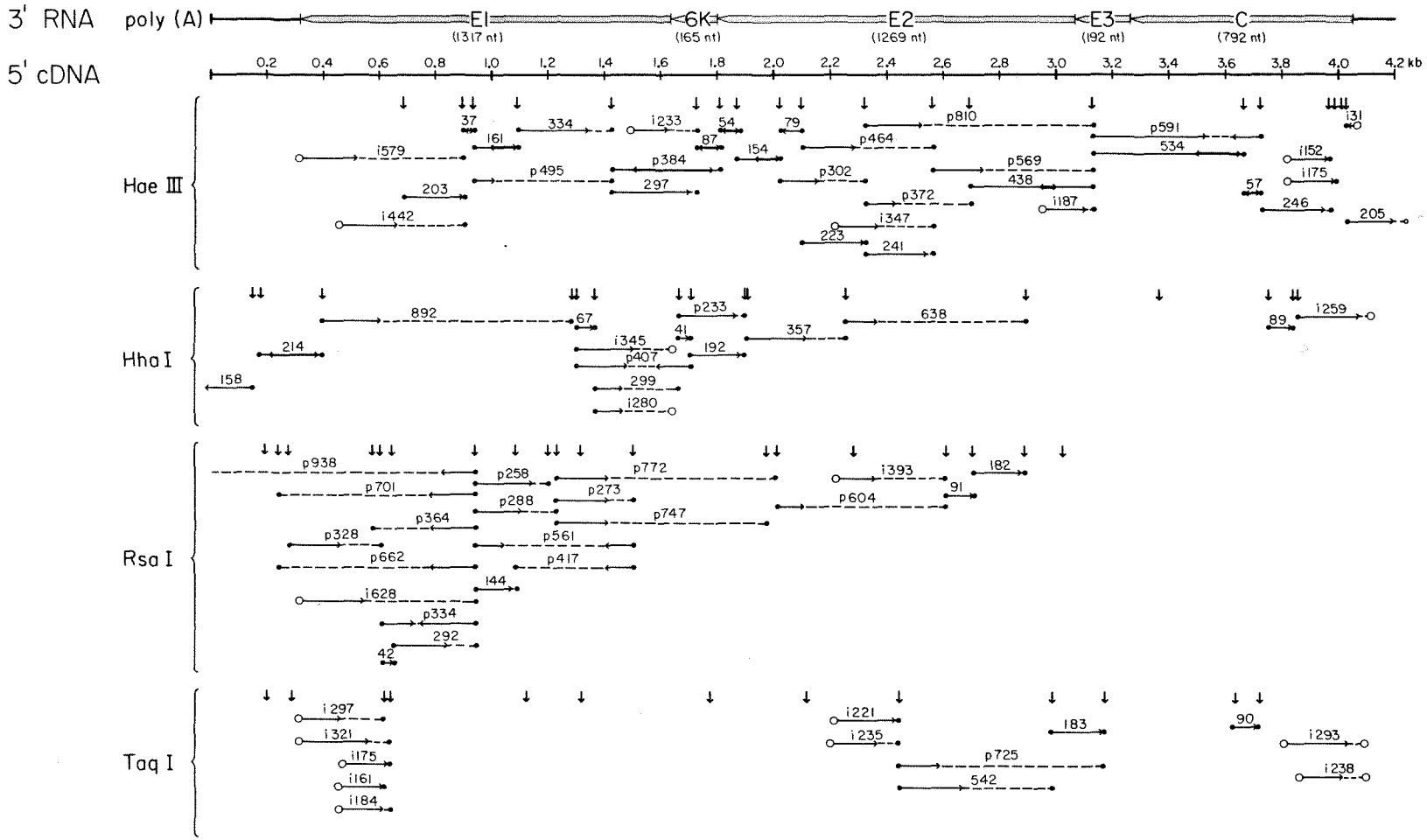

3' RNA
poly (A)
E1
(1317 nt)
6K
(165 nt)
E2
(1269 nt)
E3
(192 nt)
C
(792 nt)
5' cDNA
0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0 3.2 3.4 3.6 3.8 4.0 4.2 kb
Hae III
37 334 i233 54 79 p810 i31
161 87 p464 p591
i579 p384 154 534
p495 p302 p569 i152
203 297 438 i175
p372 57
i442 i187 246
i347 205
223 241
Hha I
892 67 p233 638 i259
214 41 357 89
i345 192
p407
158 299
i280
Rsa I
p938 p258 p772 i393 182
p701 p273 91
p288 p604
p364 p747
p328 p561
p662 p417
144
i628
p334
292
42
Taq I
i297
i321
i175
i161
i184
i221 183 90
i235 i293
p725 i238
542

Figure 6. Sequence of the cDNA complementary to Sindbis 26S RNA. The cDNA sequence is numbered from the 5' to the 3' end as described in the text. Single-stranded cDNA cleavage sites for Hae III (♦), Hha I (■), Rsa I (▼), and Taq I (●) which are the same as those found for duplex DNA are indicated by solid symbols, illegitimate sites (see text) are shown by open symbols. Areas of secondary structure causing difficulties during sequence analysis are underlined.

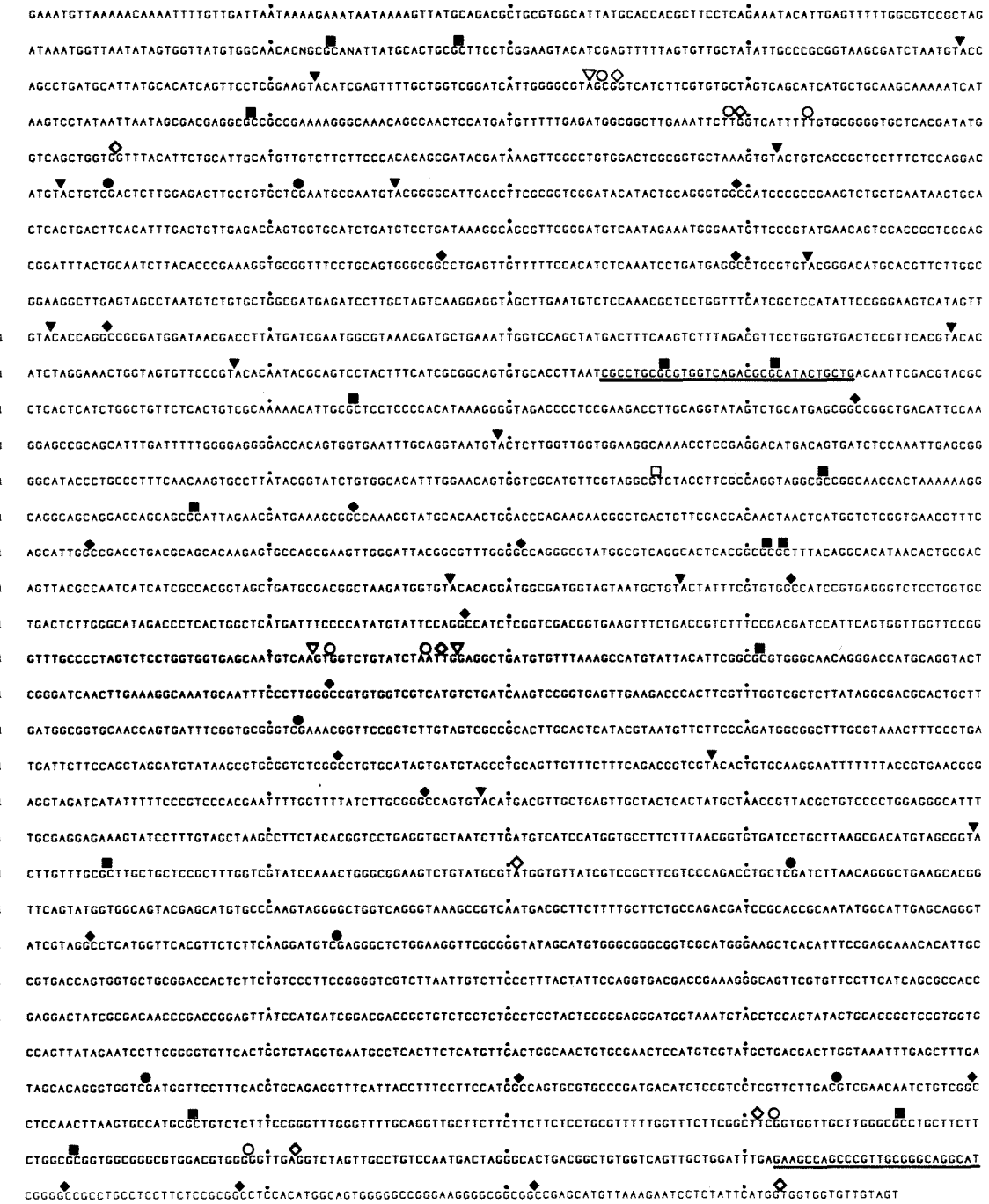


Table 1

Single-stranded cDNA restriction fragments of Sindbis 26S cDNA[a]

Enzyme	Fragment Length	Position in 26S cDNA
Hae III	p810	2319-3128
	p591	3129-3719
	i579	(314)-892
	p569	2560-3128
	534	3129-3662
	p495	930-1424
	p464	2096-2559
	i442	(451)-892
	438	2691-3128
	i403	(490)-892
	p384	1425-1808
	p372	2319-2690
	i347	(2213)-2559
	334	1091-1424
	p305	3720-4024
	p302	2017-2318
	297	1425-1721
	p268	3720-3987
	246	3720-3965
	241	2319-2559
	223	2096-2318
	i205	4025-(4229)
	203	690-892
	i187	(2942)-3128
	i175	(3814)-3987
	161	930-1090
	154	1863-2016
	i152	(3814)-3965
	i106	(2213)-2318
	i101	(3865)-3965
	87	1722-1808
	79	2017-2095
	57	3663-3719
	54	1809-1862
	37	893-929
	i31	4025-(4055)
Hha I	892	389-1280
	638	2253-2890
	p407	1295-1701
	357	1896-2252
	i345	1295-(1639)
	p340	1362-1701
	299	1362-1660
	i280	1362-(1639)

Table 1 (continued)

Enzyme	Fragment Length	Position in 26S cDNA
Hha I	i259	3847-(4105)
	p233	1661-1893
	214	175-388
	192	1702-1893
	89	3742-3830
	67	1295-1361
	41	1661-1701
Rsa I	p772	1227-2002
	p747	1227-1973
	p701	238-938
	p662	277-938
	i628	(311)-938
	p604	2003-2606
	p561	939-1499
	p417	1083-1499
	i409	(2197)-2606
	i393	(2214)-2606
	p364	575-938
	p334	605-938
	p328	277-604
	292	647-938
	p288	939-1226
	p273	1227-1499
	p258	939-1196
	182	2698-2879
	144	939-1082
	91	2607-2697
	42	605-646
Taq I	p725	2434-3158
	542	2434-2975
	i321	(313)-633
	i297	(313)-609
	i293	(3814)-(4106)
	i238	(3869)-(4106)
	i235	(2199)-2433
	i221	(2213)-2433
	i184	(450)-633
	183	2976-3158
	i175	(459)-633
	i161	(449)-609
	90	3615-3704

[a]Single-stranded cDNA restriction fragments of the 26S cDNA (Fig. 6) localized by nucleotide sequence analysis. Partial digestion products or fragments produced by illegitimate cleavage (see text) are denoted by "p" or "i", respectively, and the fragment lengths are given in nucleotides. The position of each fragment in the 26S cDNA is indicated by the 5' and 3' nucleotides. The approximate ends (±5 nucleotides) of fragments produced by illegitimate cleavage are enclosed by parentheses.

acid residue. The nucleotide sequence of the second region, in which a stable hairpin structure can be constructed, was unambiguously determined using a sequenced tryptic peptide of the capsid protein (Boege et al., 1980). In other regions causing difficulty the sequence was confirmed by reading the sequence from the 3' direction as well as the 5' direction (in which case secondary structure artifacts are found in different regions of the sequence). Occasionally we found ambiguities at certain positions in an otherwise clean sequence ladder, which did not result from the compression effects mentioned above (these ambiguous positions were present in less than 1% of the 26S cDNA sequence). This heterogeneity in the cDNA sequence may have arisen from non-random mistakes during reverse transcription, or could be due to heterogeneity in the template 26S RNA. At such positions, the predominant nucleotide is reported.

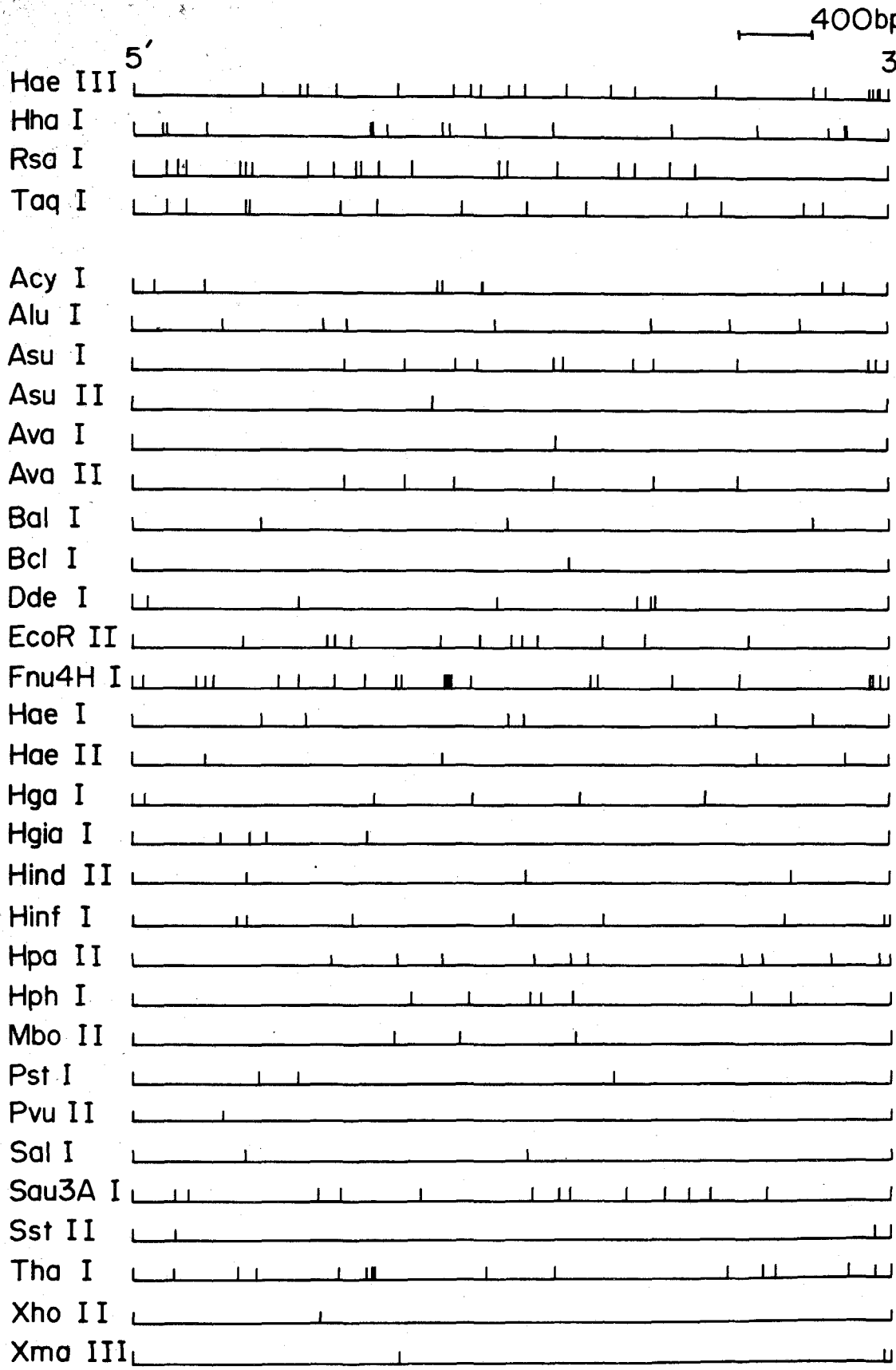
One area of the sequence near the 5' end of the cDNA (from residue 151 to 170) and located in the 3' untranslated region of the 26S RNA (J. Ou, personal communication; Rice and Strauss, 1981) was not directly determined using this sequencing method. This region has been sequenced using the dideoxy chain-termination method (Ou et al., 1981), and the unidentified nucleotides are probably due to "strong stops" for reverse transcription. A second region of the sequence near the 3' end of the cDNA corresponding to the 26S RNA sequence encoding the capsid protein could not be directly overlapped because the appropriate fragments were not isolated. The two Hae III fragments which join at nucleotide 3719 were sequenced to the end and were unambiguously aligned using the sequence of a unique tryptic peptide derived from the Sindbis capsid protein (Boege et al., 1980). It is interesting to note that although nearly all of the Hae III fragments were sequenced, several fragments from the 3' end of the cDNA were present in low yields (see Figs. 3 and 5). This implies that the RNA was not uniformly copied by reverse transcriptase. Whether this was due to the template RNA, non-random priming by the calf thymus DNA primer,

or a combination of both, is not yet known. A discussion of the validity of the overlapped 26S cDNA sequence obtained is presented in the Discussion.

As previously mentioned, the restriction pattern of 26S cDNA, digested with any of the four restriction enzymes used in this study, was more complex than expected. The direct sequence analysis of these restriction fragments has revealed several interesting properties of the cleavage of single-stranded DNA by these enzymes. Using the entire overlapped sequence (Fig. 6) we have compared the computer generated restriction patterns for duplex DNA (Fig. 7) to the actual single-stranded cDNA restriction patterns shown in Figs. 2 and 4. On the basis of size, most of the predicted fragments from a complete digestion were present in the restriction digests, which suggested that these enzymes cleave single-stranded DNA with the same specificity as double-stranded DNA (a "legitimate" cleavage site). Direct sequence analysis of all of the Hae III fragments longer than 30 nucleotides, as well as several fragments from each of the other enzymes (Fig. 5, Table 1) verified this observation. Additional fragments in Figs. 2 and 4 have been shown to be either partial digestion products or fragments probably resulting from a specific but "illegitimate" cleavage (different from the double-stranded DNA restriction site) by the restriction enzyme preparation. Although this study was by no means quantitative, it has become clear that cleavage at legitimate sites was not equivalent. Certain sites were cleaved much less efficiently than others giving rise to partial digestion products. The inability of these enzymes to cleave these partials was not due to a population of cDNA molecules which had an altered cleavage site, since sequence analysis of many of these partials revealed a normal recognition sequence at the correct position. Thus, the nonuniform restriction cleavage of single-stranded DNA by these enzymes was probably due to other aspects of the DNA structure which affected enzyme activity.

In the course of overlapping sequence data from different restriction fragments, it became apparent that some of the fragments did not originate by cleavage of

Figure 7. Computer-generated restriction map for duplex DNA complementary to Sindbis 26S RNA (see Fig. 6). The first four enzymes listed (Hae III, Hha I, Rsa I, and Taq I) cut single-stranded cDNA and were used to determine the cDNA sequence, the remaining restriction enzymes are listed in alphabetical order. The recognition sequences for the following enzymes were not found in the 26S cDNA sequence shown in Fig. 6: Aos I, BamH I, Cla I, Eca I, Eco RI, Hind III, Hpa I, Kpn I, Rsh I, Sma I, Sst I, Xba I, and Xho I.


400bp
5′
3′
Hae III
Hha I
Rsa I
Taq I
Acy I
Alu I
Asu I
Asu II
Ava I
Ava II
Bal I
Bcl I
Dde I
EcoR II
Fnu4H I
Hae I
Hae II
Hga I
Hgia I
Hind II
Hinf I
Hpa II
Hph I
Mbo II
Pst I
Pvu II
Sal I
Sau3A I
Sst II
Tha I
Xho II
Xma III

the cDNA at the expected recognition sequence. The approximate locations of these sequences are shown in Fig. 6. Most of these fragments were minor bands (see Fig. 4), implying that if they were produced by cleavage these sites were cut less efficiently than legitimate recognition sites. Contaminating specific endonucleases in these preparations could have been responsible for these cleavages, although inspection of nucleotide sequences in the vicinity of these cleavage sites failed to reveal any obvious pattern of additional specific recognition sequences. It is of interest to note that several of the illegitimate sites are shared by two or more enzyme preparations (Fig. 6). Even though additional experiments are needed to clarify the origin of these fragments, they have proven extremely useful for 26S cDNA sequence analysis.

Base composition, nearest-neighbor analysis, and codon usage. From the sequence presented in Fig. 6, the base composition for this portion of the 26S RNA is 24%G, 28%C, 27%A, and 21%U. Table 2 presents the computer generated nearest-neighbor analysis for this sequence. Although some of the doublet frequencies appear to be non-random, we do not find a scarcity of the CpG doublet as has been found for several vertebrate DNAs (Russell et al., 1976; Bird, 1980) and several RNA(Porter et al., 1979; Jou et al., 1980) and DNA (Fiers et al., 1978; Soeda et al., 1980) viruses with vertebrate hosts. In addition, the frequency with which duplex DNA complementary to the 26S RNA would be cleaved by various restriction enzymes (Fig. 7) is approximately that expected based on the length of their recognition sequences.

Table 3 shows the codon usage in the coding resions of Sindbis virus 26S RNA. The distribution among the possible codons for most of the amino acids is clearly non-random (see for example, leucine and arginine), and would not be predicted using the base composition (see above) or nearest-neighbor frequencies (Table 2) found for the 26S RNA. In addition, in spite of the extensive nucleotide sequence divergence between Sindbis virus and SFV (see Fig. 8 and Discussion) the preference for certain codons is remarkably similar (Table 3).

Table 2

Sindbis 26S RNA Nearest Neighbor Frequencies[a]

	U	A	G	C
U	197	175	260	209
A	234	299	270	307
G	203	279	217	296
C	207	356	247	309

[a]Doublet frequencies (5' bases are in the left margin) are for the complement of the 26S cDNA sequence shown in Fig. 6. The first figure is the number of times the doublet is found in the sequence (4065 doublets were examined).

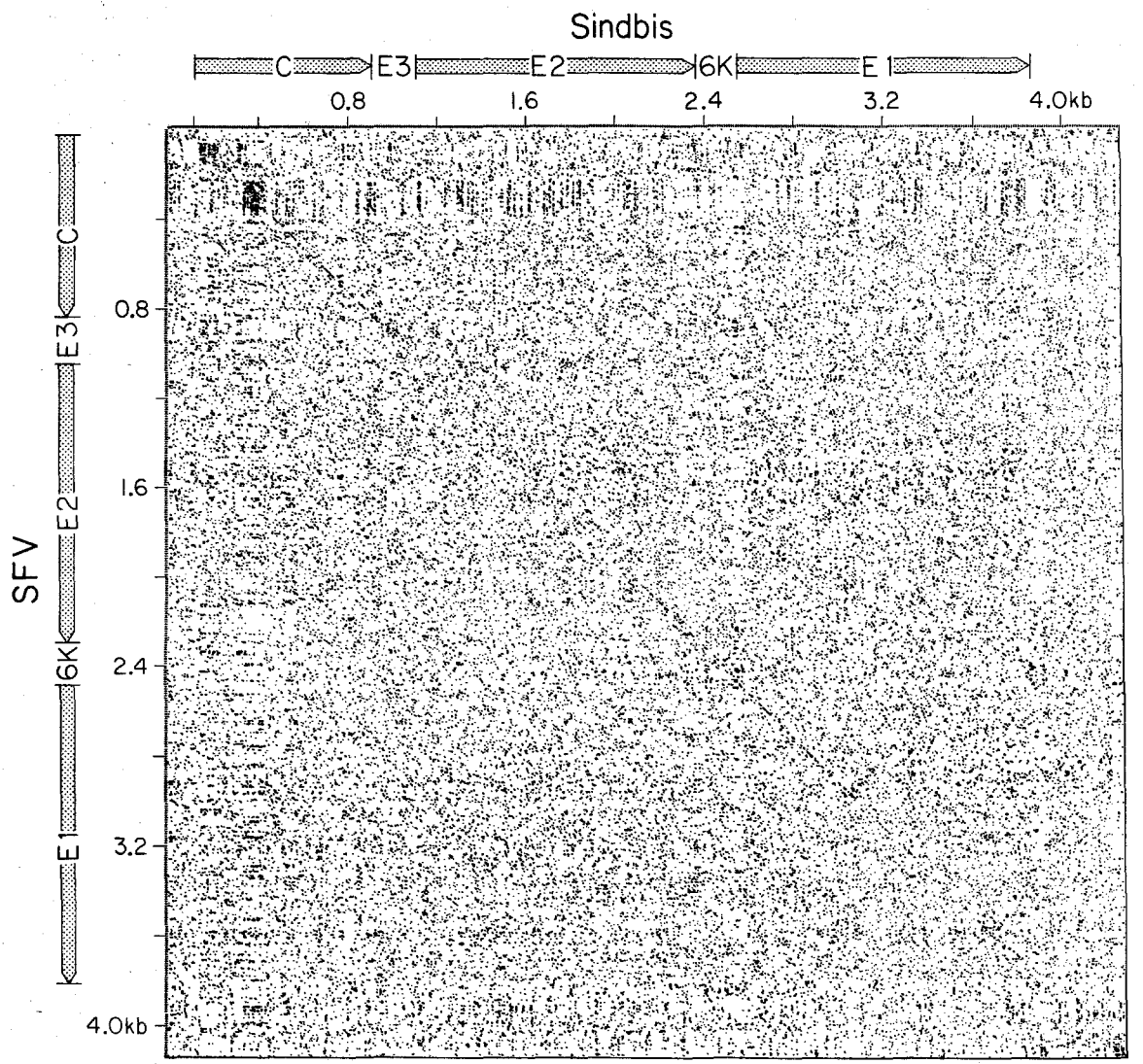
Table 3

Use of Codons in Sindbis and Semliki Forest Virus 26S RNAs.[a]

PHE	UUU	19	10	SER	UCU	6	9	TYR	UAU	15	16	CYS	UGU	12	14
	UUC	24	29		UCC	13	10		UAC	31	35		UGC	31	34
LEU	UUA	5	4		UCA	21	14	OCH	UAA	0	1	OPL	UGA	1	0
	UUG	22	14		UCG	17	21	AMB	UAG	0	0	TRP	UGG	15	17
LEU	CUU	8	8	PRO	CCU	20	20	HIS	CAU	18	17	ARG	CGU	5	2
	CUC	16	13		CCC	14	15		CAC	21	27		CGC	19	13
	CUA	8	10		CCA	30	19	GLN	CAA	17	18		CGA	4	2
	CUG	25	26		CCG	28	35		CAG	25	31		CGG	4	7
ILE	AUU	19	12	THR	ACU	20	17	ASN	AAU	12	12	SER	AGU	15	7
	AUC	27	28		ACC	41	28		AAC	28	37		AGC	23	15
	AUA	13	15		ACA	23	28	LYS	AAA	35	32	ARG	AGA	17	17
MET	AUG	29	28		ACG	14	24		AAG	42	48		AGG	11	14
VAL	GUU	14	13	ALA	GCU	12	23	ASP	GAU	14	11	GLY	GGU	7	8
	GUC	36	34		GCC	41	40		GAC	37	43		GGC	19	23
	GUA	16	16		GCA	33	25	GLU	GAA	31	26		GGA	38	24
	GUG	24	41		GCG	22	23		GAG	23	25		GGG	16	26

[a]The first figure after each codon is the number of times that codon is used in the translated regions of Sindbis 26S RNA (deduced from the cDNA sequence in Fig. 6; Rice and Strauss, 1981). The second number is the corresponding figure for SFV (from Garoff et al., 1980a,b).

Figure 8. Nucleotide sequence homology between the 26S RNAs of Sindbis virus and SFV. The SFV RNA sequence data is taken from Garoff et al. (1980a,b). A dot in the matrix represents a 5 out of 6 nucleotide identity between the two sequences. The diagonal line shows that the two sequences are homologous, but contain numerous regions with little nucleotide sequence homology. The dense cluster in the upper left hand corner is caused by a region of the sequence rich in A, G, and C but lacking U.


Sindbis
C
E3
E2
6K
E 1
0.8
1.6
2.4
3.2
4.0kb
SFV
C
E3
E2
6K
E1
0.8
1.6
2.4
3.2
4.0kb

DISCUSSION

Techniques for rapid sequence analysis have been developed to determine the primary structure of both RNA (Donis-Keller et al., 1977; Simoncsitis et al., 1977) and DNA (Sanger et al., 1977; Maxam and Gilbert, 1980). Molecular cloning has enabled the amplification of rare cellular mRNA sequences such that primary sequence data can be obtained (for example, see Taniguchi et al., 1980; Steinmetz et al., 1981). However, many RNA viruses produce large quantities of virus-specific RNA in infected cells, and package single-stranded genomic RNA into mature virions which can be readily purified. In such cases, molecular cloning is not necessary for nucleotide sequence determination. While this work was in progress, Kitmura and Wimmer (1980) published extensive sequence data on poliovirus without the use of molecular cloning. Their ingenious approach involved the chain-termination sequencing method (Sanger et al., 1977) using RNase resistant oligonucleotides to prime DNA synthesis on a cDNA template made from poliovirus RNA. This method uses relatively large amounts of viral RNA (>100 μg), however, and for this reason would not be suitable for analysis of viruses which grow poorly, such as many ts mutants of Sindbis virus. Other methods include direct sequence analysis of restriction fragment primed cDNA (Ghosh et al., 1980), or isolation and sequence analysis of 5' end-labeled ribonuclease T_1-resistant oligonucleotides (Pederson and Haseltine, 1980). Alternatively, the method we have used for determining the Sindbis 26S RNA sequence requires only a few micrograms of cDNA, and we estimate that most of the genomic 49S RNA could be sequenced with less than 5-10 μg of cDNA. This sequencing strategy, without utilizing prior restriction mapping, is similar to the "shotgun" approach in M13 used by Sanger et al. (1980) for the human mitochondrial genome, and also for influenza virus (Winter and Fields, 1980). Randomly primed cDNA synthesis gives a complete representation of the RNA sequences which can then be digested by a growing number

of Type II restriction endonucleases which cleave single-stranded DNA into discrete fragments (Blakesley and Wells, 1975; Horiuchi and Zinder, 1975; Godson and Roberts, 1976; Molloy and Symons, 1980). Unlike other procedures, this mixture of single-stranded DNA fragments can then be end-labeled, separated on sequencing gels and sequenced using the base-specific chemical cleavage method of Maxam and Gilbert (1980) without the secondary restriction enzyme digestion or strand separation required for sequence analysis of end-labeled duplex DNA restriction fragments. 3' terminal RNA sequences of poly(A) containing RNA can be determined if an oligo(dT) primer is used for cDNA synthesis, and the restriction enzyme cleaved cDNA purified by passage over an oligo(dA) cellulose column after 3' end-labeling. This is also applicable to RNAs which lack poly(A) if polyadenylation can be performed in vitro (Emtage et al., 1979). Recent improvements in the base-specific chemical cleavage method (Smith and Calvo, 1980), have allowed sequences in excess of 600 bases from one labeled end to be determined. Since either end of a fragment can be labeled, this extends the fragment length for which complete preliminary sequence data can be obtained to 1.2 kb. Thus, using restriction enzymes which cleave less frequently (those with 5 and 6 base recognition sequences), it should be possible to rapidly sequence even very large (>10 kb) viral RNAs.

cDNA synthesis. The optimal conditions for production of approximately full length (4.2 kb) cDNA to intracellular Sindbis 26S RNA are similar to those reported for other RNAs (Kacian and Myers, 1976; Buell et al., 1978; Zain et al., 1979; Devos et al., 1979). However, the pattern of oligo(dT) primed cDNA products on alkaline agarose gels reveals numerous smaller discrete bands. Such products have been found in other systems (Owens and Cress, 1980) and may be due to RNA secondary structure or internal priming, although several other explanations are possible (Kacian and Meyers, 1976). The similarity in the pattern of oligo(dT) primed cDNA products synthesized from 26S and 49S Sindbis RNAs, strongly suggests that internal priming

is not responsible for the production of most of these smaller species. There is an apparent "strong stop" for reverse transcriptase approximately 4.2 kb from the 49S RNA poly(A) tail, which could reflect an RNA structural feature important in the regulation of 26S and 49S transcription. Prior denaturation of template RNA with methylmercury hydroxide has been found to improve both the yield and length of cDNA for certain mRNAs (Payvar and Schimke, 1979). This method may prove useful for the generation of longer cDNA transcripts of the 49S RNA.

cDNA cleavage. The ability to specifically cleave this single-stranded cDNA with Hae III and Hha I (Blakesley and Wells, 1975; Godson and Roberts, 1976), in addition to Rsa I and Taq I, has facilitated its complete sequence determination as well as enabling us to compare virion and intracellular RNAs. In the case of the 26S cDNA, more than 90% of the Hae III fragments were virus-specific as determined by comparison to the 49S cDNA Hae III restriction pattern, which was later verified by direct sequence analysis. The extra minor fragments are probably due to contamination of the Sindbis 26S RNA by poly(A) containing cellular RNA or residual rRNA not completely removed by oligo(dT) cellulose chromatography. The molar yield of these nonviral Hae III fragments was usually less than 10% of the yield for virus-specific fragments. After determining the cDNA sequence, it was possible, with the aid of a computer program, to generate a complete restriction map for double-stranded DNA. The comparison of this map with the actual restriction digests, in addition to sequence analysis of many of the fragments, strongly suggests that Hae III (see also: Horiuchi and Zinder, 1975; Blakesley and Wells, 1975), Hha I, Rsa I and Taq I recognize and cleave the same sequences in single-stranded DNA as double-stranded DNA. Many of the other fragments were partial digestion products, but some were apparently produced by cleavage at illegitimate sites. These fragments produced by illegitimate cleavage were usually minor species. Other investigators

have observed quantitative differences in the sensitivity of Hae III sites in single-stranded DNA bacteriophage f1 (Horiuchi and Zinder, 1975), and additional fragments in Hae III digestion of $\phi\chi174$ virion single-stranded DNA (Blakesley and Wells, 1975). It is of note that several of the enzyme preparations used in this study cleave at similar illegitimate sites in the 26S cDNA. However, it is not known whether contaminating endonucleases or additional specificities of these restriction endonucleases are responsible for these cleavages. In either case, these fragments are useful for sequence analysis, but one should exercise caution in inferring double-stranded DNA restriction sites from single-stranded DNA restriction patterns. The rigorous purification and characterization of these and other endonucleases which specifically cleave single-stranded DNA could contribute greatly to the rapid sequence analysis and comparative study of viral and other readily purified RNAs.

Validity of the cDNA sequence. The accuracy of the 26S cDNA sequence is supported by several lines of evidence. Translation of the complement to the 26S cDNA sequence has revealed one open reading frame of 3735 nucleotides beginning with an AUG codon. Sindbis 26S RNA is known to have a single initiation site for protein synthesis (Cancedda et al., 1975), and the precursors of the structural proteins are translated in the order C, PE2 (precursor to E3 and E2), and E1 (Schlesinger and Kääriäinen, 1980). Our deduced polypeptide sequence is in complete agreement with this gene order and with the NH_2-terminal sequences found by automated sequence analysis of E1 and E2 (Bell et al., 1978), of E3 and PE2 (Bonatti and Blobel, 1979; Bell et al., manuscript in preparation; Mayne et al., manuscript in preparation), of the 6K protein (Welch and Sefton, 1979; B. Sefton, personal communication), and of the hydrophobic transmembrane "roots" of E1 and E2 derived by treatment of intact virions with α-chymotrypsin (Rice et al., manuscript in preparation). In addition, Boege et al. (1980) have recently published an extensive study on tryptic

peptides from the Sindbis (strain Sa-AR-86) capsid protein. Most of this sequence and composition data are consistent with our deduced sequence for the capsid protein. The amino acid compositions of the structural proteins determined by amino acid analysis (Bell et al., 1979) are in excellent agreement with the compositions determined from the nucleotide sequence data (Table 4). In addition, our sequence data adjacent to the poly(A) tail is in complete agreement with the sequence obtained for the same region using a chain termination method (Ou et al., 1981).

We have recently compared the deduced protein sequences encoded by Sindbis 26S RNA (Rice and Strauss, 1981) and those of a closely related alphavirus, Semliki Forest virus (Garoff et al., 1980a,b). There is striking protein sequence homology (47% of the residues are identical, and another 12% are conservative substitutions) and very similar codon utilization (Table 3). The common preference for certain codons is surprising in view of the extensive nucleotide sequence divergence between Sindbis and SFV (see below and Fig. 8). In addition, it is of interest to note that although vertebrates are among the natural hosts for these viruses, they do not have the low frequency of the CpG doublet found in vertebrate DNAs (Russell et al., 1976; Bird, 1980) and both DNA (Fiers et al., 1978; Soeda et al., 1980) and RNA viruses (Porter et al., 1979; Jou et al., 1980) with vertebrate hosts. The similar codon utilization of Sindbis and SFV may reflect a common adaptive constraint for optimal translational efficiency in their wide host range which includes insects (whose DNAs do not contain low CpG frequencies; see Bird, 1980) as well as birds and mammals.

The 26S RNA nucleotide sequences of Sindbis virus and SFV are compared in Fig. 8. The strong diagonal in Fig. 8 demonstrates that both Sindbis virus and SFV have evolved from a common ancestor largely by base substitution with relatively few deletions or insertions in the nucleic acid sequence encoding their structural proteins. This homology breaks down in most of the 3' untranslated region and presumably reflects less stringent evolutionary constraints on these sequences. In addition,

Table 4

Amino Acid Compositions of the Three Structural Proteins of Sindbis Virus[a]

	CAPSID		E2		E1	
	from sequence	from analyzer	from sequence	from analyzer	from sequence	from analyzer
ASP	9		21		17	
	} 17		} 33		} 34	
ASN	8		11		16	
THR	17	17	40	40	34	34
SER	12	12	30	30	43	43
GLU	12		17		19	
	} 26		} 30		} 35	
GLN	14		11		15	
PRO	28	28	29	30	27	28
GLY	24	24	25	27	27	28
ALA	22	22	32	33	42	42
CYS	0	0	17	14	17	15
VAL	15	15	31	31	36	35
MET	10	10	6	6	10	10
ILE	8	8	23	22	24	23
LEU	14	14	28	29	27	28
TYR	4	4	22	18	16	13
PHE	9	9	9	10	19	19
HIS	6	6	19	17	13	13
LYS	25	25	27	27	23	23
ARG	23	22	20	19	10	10
TRP	4	4	5	6	4	4
TOTAL	264		423		439	
MOL. WT.	29,322		46,835		47,301	

Table 4 (continued)

[a]Sequence composition is deduced from the complement of the cDNA sequence in Fig. 6 (Rice and Strauss, 1981). Analyzer composition is from Bell et al., 1979; data have been rounded to the nearest integer assuming 264, 423, and 439 residues in C, E2, E1, respectively. Molecular weights are for the peptide part only.

there are nucleotide sequences which encode protein which are much less conserved than others. For example, the NH_2-terminal region of the Sindbis and SFV capsid protein shows very little protein or nucleic acid sequence homology, but in both viruses contains a clustering of lysine, arginine, and proline forming a highly basic domain (Garoff et al., 1980b; Rice and Strauss, 1981). In addition, the nucleotide sequences encoding the E2 glycoproteins show very little homology except near the COOH termini of the glycoproteins. The protein and nucleic acid sequence homology between these two closely related alphaviruses provides additional support for the validity of the sequence data obtained by this method.

ACKNOWLEDGEMENTS

We would like to thank E. Lenches for preparing the chick fibroblasts, T. E. Crowley for help in the early stages of this work, H. V. Huang, T. Hunkapiller, and M. and R. Douglas for writing computer programs and J. R. Bell for the preparation of some of the figures, and L. Hood and T. Maniatis for their generosity in lending us supplies and equipment. This work was supported by Grants PCM 80-22830 from the National Science Foundation, by Grants GM 06965 and AI 10793 from NIH, and by Biomedical Research Support Grant 507 RR 07003 from NIH. CMR was supported by Training Grant GM 00086 from NIH.

REFERENCES

Bell, J. R., Hunkapiller, M. W., Hood, L. E., and Strauss, J. H. (1978) Aminoterminal sequence analysis of the structural proteins of Sindbis virus. Proc. Natl. Acad. Sci. USA **75**, 2722-2726.

Bell, J. R., Strauss, E. G., and Strauss, J. H. (1979) Purification and amino acid compositions of the structural proteins of Sindbis virus. Virology **97**, 287-294.

Berger, S. L., and Birkenmeier, C. S. (1979) Inhibition of intractable nuclease with ribonucleoside-vanadyl complexes: isolation of messenger ribonucleic acid from resting lymphocytes. Biochemistry **18**, 5143-5149.

Bird, A. P. (1980) DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. **8**, 1499-1504.

Blakesley, R. W., and Wells, R. D. (1975) 'Single-stranded' DNA from $\phi\chi174$ and M13 is cleaved by certain restriction endonucleases. Nature (London) **257**, 421-422.

Boege, U., Wengler, G., Wengler, G., and Wittman-Liebold, B. (1980) Partial amino acid sequences of Sindbis and Semliki Forest virus-specific core proteins. Virology **103**, 178-190.

Bonatti, S., and Blobel, G. (1979) Absence of a cleavable signal sequence in Sindbis virus glycoprotein PE2. J. Biol. Chem. **254**, 12261-12264.

Buell, G. N., Wickens, M. P., Payvar, F., and Schimke, R. T. (1978) Synthesis of full length cDNAs from four partially purified oviduct mRNAs. J. Biol. Chem. **253**, 2471-2482.

Burge, B. W., and Pfefferkorn, E. R. (1966) Isolation and characterization of conditional-lethal mutants of Sindbis virus. Virology **30**, 204-213.

Cancedda, R., Villa-Komaroff, L., Lodish, H. F., and Schlesinger, M. (1975) Initiation sites for translation of Sindbis virus. Cell **6**, 215-222.

Chang, J. C., Temple, G. F., Poon, R., Neumann, K. H., and Kan, Y. W. (1977) The nucleotide sequences of the untranslated 5' regions of human α- and β-globin mRNAs. Proc. Natl. Acad. Sci. USA **74**, 5145-5149.

Devos, R., van Emmelo, J., Contreras, R., and Fiers, W. (1979) Construction and characterization of a plasmid containing a nearly full-size DNA copy of bacteriophage MS2 RNA. J. Mol. Biol. **128**, 595-619.

Donis-Keller, H., Maxam, A. M., and Gilbert, W. (1977) Mapping adenines, guanines and pyrimidines in RNA. Nucleic Acids Res. **4**, 2527-2538.

Emtage, J. S., Catlin, G. H., and Carey, N. H. (1979) Polyadenylation and reverse transcription of influenza viral RNA. Nucleic Acids Res. **6**, 1221-1239.

Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G., and Ysebaert, M. (1978) Complete nucleotide sequence of SV40 DNA. Nature (London) **273**, 113-120.

Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H., and Delius, A. (1980a) Nucleotide sequence of cDNA coding for Semliki Forest membrane glycoproteins. Nature (London) **288**, 236-241.

Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H., and Delius, H. (1980b) The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino terminal region. Proc. Natl. Acad. Sci. USA **77**, 6376-6380.

Ghosh, P. K., Reddy, V. B., Piatak, M., Lebowitz, P., and Weissman, S. M. (1980) Determination of RNA sequences by primer directed synthesis and sequencing of their cDNA transcripts. in Methods in Enzymology, Grossman, L., and Moldave, K., eds. (Academic Press, New York) Vol. 65, pp. 580-595.

Godson, G. N., and Roberts, R. J. (1976) A catalogue of cleavages of ϕχ174, S13, G4, and ST-1 DNA by 26 different restriction endonucleases. Virology **73**, 561-567.

Horiuchi, K., and Zinder, N. D. (1975) Site-specific cleavage of single-stranded DNA by a Haemophilus restriction endonuclease. Proc. Natl. Acad. Sci. USA **72**, 2555-2558.

Hsu, M. T., Kung, H. J., and Davidson, N. (1973) An electron microscope study of Sindbis virus RNA. Cold Spring Harbor Symp. Quant. Biol. **38**, 943-950.

Jou, W. M., Verhoeyen, M., Devon, R., Saman, E., Fang, R., Huylebroeck, D., Fiers, W., Threlfall, G., Barber, C., Carey, N., and Emtage, S. (1980) Complete structure of the hemagglutinin gene from the human influenza A/Victoria/3/75 (H3N2) strain as determined from cloned DNA. Cell **19**, 683-696.

Kacian, D. L., and Myers, J. C. (1976) Synthesis of extensive, possibly complete, DNA copies of poliovirus RNA in high yields at high specific activities. Proc. Natl. Acad. Sci. USA **73**, 2191-2195.

Kennedy, S. I. T. (1976) Sequence relationships between the genome and the intracellular RNA species of standard and defective-interfering Semliki Forest virus. J. Mol. Biol. **108**, 491-511.

Kitmura, N., and Wimmer, E. (1980) Sequence of 1060 3'-terminal nucleotides of poliovirus RNA as determined by a modification of the dideoxynucleotide method. Proc. Natl. Acad. Sci. USA **77**, 3196-3200.

Laskey, R. A., and Mills, A. D. (1977) Enhanced autoradiographic detection of ^{32}P and ^{125}I using intensifying screens and hypersensitized film. FEBS Lett. **82**, 314-316.

Maxam, A. M., and Gilbert, W. (1980) "Sequencing with base-specific chemical cleavages." In Methods in Enzymology, Grossman, L., and Moldave, K., eds. (Academic Press, New York) Vol. 65, pp. 499-560.

McDonell, M. W., Simon, M. N., and Studier, F. W. (1977) Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis on neutral and alkaline gels. J. Mol. Biol. **110**, 119-146.

Molloy, P. L., and Symons, R. H. (1980) Cleavage of DNA·RNA hybrids by Type II restriction endonucleases. Nucleic Acids Res. **8**, 2939-2946.

Myers, J. C., and Spiegelman, S. (1978) Sodium pyrophosphate inhibition of RNA·DNA hybrid degradation by reverse transcriptase. Proc. Natl. Acad. Sci. USA **75**, 5329-5333.

Ou, J.-H., Strauss, E. G., and Strauss, J. H. (1981) Comparative studies of the 3'-terminal sequences of several alphavirus RNAs. Virology, in press.

Owens, R. A., and Cress, D. E. (1980) Molecular cloning and characterization of potato spindle tuber viroid cDNA sequences. Proc. Natl. Acad. Sci. USA **77**, 5302-5306.

Payvar, F., and Schimke, R. T. (1979) Methylmercury hydroxide enhancement of translation and transcription of ovalbumin and conalbumin mRNAs. J. Biol. Chem. **254**, 7636-7642.

Pederson, F. S., and Haseltine, W. A. (1980) "A micromethod for detailed characterization of high molecular weight RNA" in Methods in Enzymology. Grossman, L., and Moldave, K., eds. (Academic Press, New York) Vol. 65, pp. 680-687.

Pierce, J. S., Strauss, E. G., and Strauss, J. H. (1974) Effect of ionic strength on the binding of Sindbis virus to chick cells. J. Virol. **13**, 1030-1036.

Porter, A.G., Barber, C., Carey, N. H., Hallewell, R. A., Threlfall, G., and Emtage, J. S. (1979) Complete nucleotide sequence of an influenza virus hemagglutinin gene from cloned DNA. Nature (London) **282**, 471-477.

Rice, C. M., and Strauss, J. H. (1981) Nucleotide sequence of the 26S mRNA of Sindbis virus and deduced sequence of the encoded virus structural proteins. Proc. Natl. Acad. Sci. USA, in press.

Russell, G. J., Walker, P. M. B., Elton, R. A., and Subak-Sharpe, J. H. (1976) Doublet frequency analysis of fractionated vertebrate nuclear DNA. J. Mol. Biol. **108**, 1-23.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**, 161-178.

Sanger, F., Nicklen, S., and Coulson, A. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**, 5463-5467.

Sato, S., Hutchison, C. A., III, and Harris, J. I. (1977) A thermostable sequence-specific endonuclease from Thermus aquaticus. Proc. Natl. Acad. Sci. USA **74**, 542-546.

Schlesinger, M. J., and Kääriäinen, L. (1980) "Translation and processing of alphavirus proteins" in The Togaviruses, Schlesinger, R. W., ed. (Academic Press, New York) pp. 371-392.

Simmons, D. T., and Strauss, J. H. (1972) Replication of Sindbis virus. I. Relative size and genetic content of 26S and 49S RNA. J. Mol. Biol. **71**, 599-613.

Simoncsitis, A., Brownlee, G. G., Brown, R. S., Rubin, J. R., and Guilley, H. (1977) New rapid gel sequencing method for RNA. Nature (London) **269**, 833-836.

Smith, D. R., and Calvo, J. M. (1980) Nucleotide sequence of the E. coli gene coding for dihydrofolate reductase. Nucleic Acids Res. **8**, 2255-2274.

Soeda, E., Arrand, J. R., and Griffin, B. E. (1980) Polyomavirus DNA: Complete nucleotide sequence of the gene which codes for polyoma virus capsid protein VP1 and overlaps the VP2/VP3 genes. J. Virol. **33**, 619-630.

Steinmetz, M., Frelinger, J. B., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S., and Hood, L. (1981) Characterization of three cDNA clones encoding mouse transplantation antigens: homology to immunoglobulins. Cell, in press.

Strauss, E. G., and Strauss, J. H. (1980) "Mutants of alphaviruses: genetics and physiology" in the Togaviruses, Schlesinger, R. W., ed. (Academic Press, New York) pp. 393-426.

Strauss, E. G. (1978) Mutants of Sindbis virus. III. Host polypeptides present in purified HR and ts 103 virus particles. J. Virol. **28**, 466-474.

Strauss, E. G., and Strauss, J. H. (1977) "Togaviruses" in The Molecular Biology of Animal Viruses, Nayak, D., ed. (Dekkor, New York) Vol. 1, pp. 111-166.

Taniguchi, T., Fujii-Kuriyama, Y., and Muramatsu, M. (1980) Molecular cloning of human interferon cDNA. Proc. Natl. Acad. Sci. USA **77**, 4003-4006.

Welch, W. J., and Sefton, B. M. (1979) Two small virus-specific polypeptides are produced during infection with Sindbis virus. J. Virol. **29**, 1186-1195.

Wengler, G. (1980) "Effects of alphaviruses on host cell macromolecular synthesis." in The Togaviruses, Schlesinger, R. W., ed. (Academic Press, New York) pp. 459-472.

Winter, G., and Fields, S. (1980) Cloning of influenza cDNA into M13: the sequence of the RNA segment encoding the A/PR/8/34 matrix protein. Nucleic Acids Res. **9**, 1965-1974.

Zain, S., Sambrook, J., Roberts, R. J., Keller, W., Fried, M., and Dunn, A. R. (1979) Nucleotide sequence analysis of the leader segments in a cloned copy of Adenovirus 2 fiber mRNA. Cell **16**, 851-861.

CHAPTER 3

Nucleotide sequence of the 26S mRNA of Sindbis virus and deduced sequence of the encoded virus structural proteins

CHARLES M. RICE and JAMES H. STRAUSS

Proceedings of the National Academy of Sciences, in press.

ABSTRACT The nucleotide sequence of intracellular 26S mRNA of Sindbis virus has been determined by direct sequence analysis of the cDNA made to this RNA with reverse transcriptase. From this study, the amino acid sequences of the encoded virus structural proteins, which include a basic capsid protein and two integral membrane glycoproteins, have been deduced. The features of these proteins as related to their functions are discussed. We suggest that 3 proteases are required to produce these proteins from their polyprotein precursor: a viral protease functions in the cytosol to release the capsid protein; signalase makes two cleavages which separate the glycoproteins; and a protease of the Golgi complex which cleaves after double basic residues processes the precursor form of one of the glycoproteins.

INTRODUCTION

Alphaviruses such as Sindbis virus and the closely related Semliki Forest virus are simple enveloped viruses. The icosahedral nucleocapsid is assembled in the cytoplasm, diffuses to the cell surface and buds through the host cell plasmalemma, acquiring a lipoprotein envelope containing only virus encoded glycoproteins. These glycoproteins are synthesized on the rough endoplasmic reticulum, glycosylated, and migrate to the plasma membrane by way of the Golgi apparatus, where the carbohydrates are modified and lipids are covalently attached. The interaction between the alphavirus nucleocapsid and its glycoproteins is much more specific than in other enveloped viruses; mature virions contain exclusively alphavirus proteins (1).

All three of the virus structural proteins are translated as a continuous polypeptide from a single messenger RNA molecule, called 26S RNA (2). We have determined the nucleotide sequence of Sindbis HR 26S RNA in order to investigate in detail the structure and processing of the viral proteins and to make possible further study of the temperature sensitive mutants previously characterized genetically and physiologically (3).

MATERIALS AND METHODS

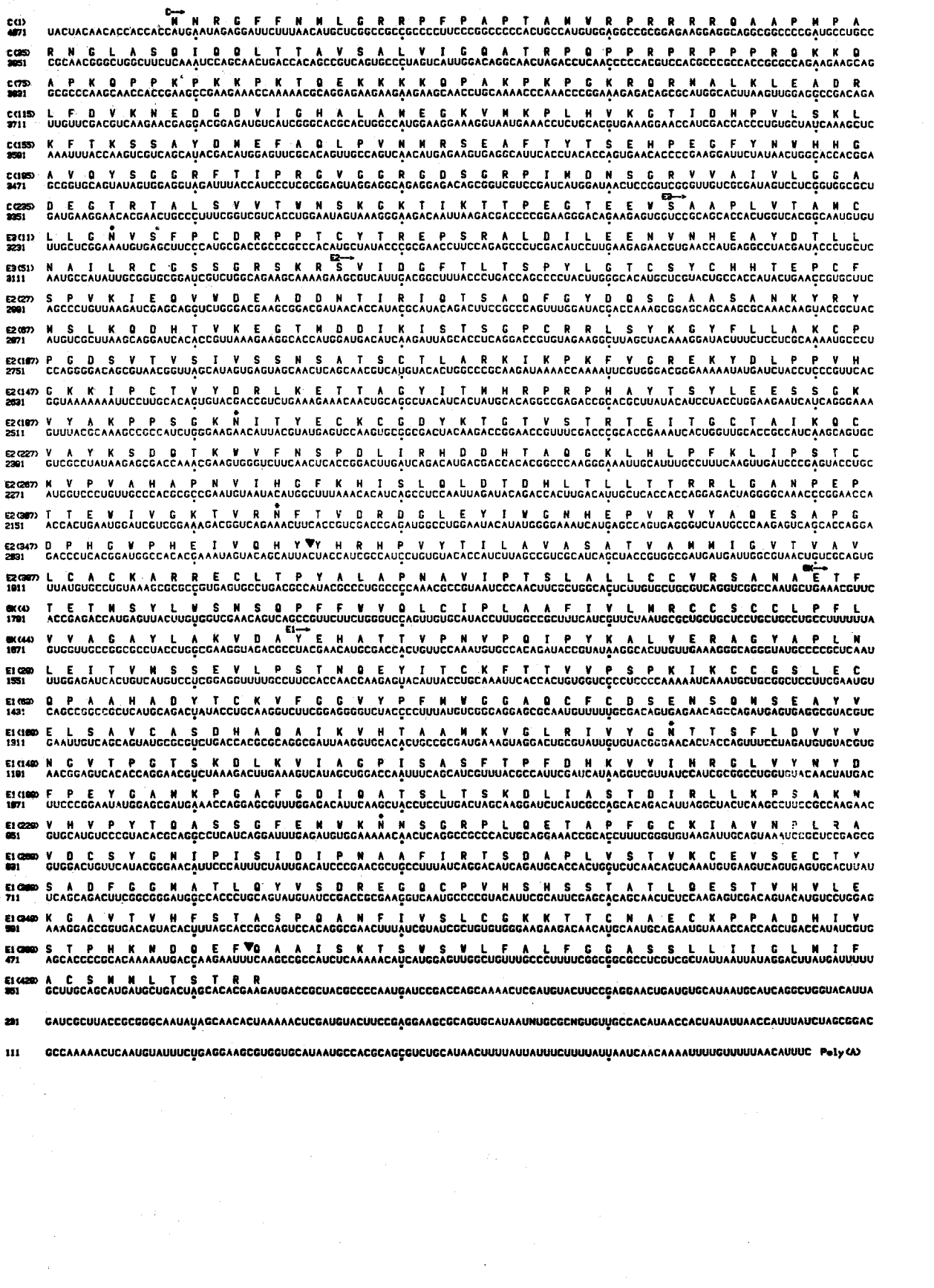
Details of the methods and strategy used for preparation and sequencing of single-stranded cDNA to Sindbis 26S RNA will be published elsewhere. Briefly, 26S RNA was used as a template for synthesis of complementary DNA (cDNA) at 42.5°C using avian myeloblastosis virus reverse transcriptase (kindly provided by J. Beard) and primed with either oligo$(dT)_{12-18}$ (Collaborative Research) or a mixture of short (~6-8 nucleotides) random oligonucleotides derived from calf thymus DNA (a gift from J. Casey). The reaction mixture contained 4 mM sodium pyrophosphate to inhibit second strand synthesis. After 30-60 minutes an excess of Na_2 EDTA was added to stop the reaction followed by phenol/chloroform extraction, and ethanol

precipitation of the cDNA-RNA hybrid. The RNA strand was hydrolyzed by incubation in 0.1 M NaOH at 60°C for 30 minutes. The cDNA was chromatographed over Biogel A5M (BioRad Labs); the excluded peak fractions were pooled and ethanol precipitated. The cDNA after digestion with either Hae III, Taq 1, Hha 1, or Rsa 1 (New England Biolabs) was labeled at the 5' ends with T4 polynucleotide kinase or at the 3' ends with terminal deoxynucleotidyl transferase (Bethesda Research Labs, ribosubstitution grade) essentially as described by Maxam and Gilbert (4). Single-stranded, end-labeled restriction fragments were separated on 6% polyacrylamide sequencing gels, excised, eluted, and sequenced using the base specific chemical cleavage procedure (4). The modifications of Smith and Calvo (5) were employed for long (>300 nucleotides) fragments. By using four different restriction enzymes, sequencing numerous partial digestion products, and sequencing some fragments from both the 5' and the 3' direction, virtually all of the sequence was determined more than once and sufficient overlap was obtained to align all of the fragments.

RESULTS AND DISCUSSION

Sequence of 26S RNA. Figure 1 presents the entire nucleotide sequence of Sindbis virus (HR strain) 26S RNA, excluding a sequence of about 150 nucleotides at the 5' end which has not been unambiguously determined, and the deduced amino acid sequences of the encoded proteins. From the AUG codon initiating synthesis of the capsid protein, there is an open reading frame for 3735 nucleotides encoding, in order, capsid protein, E3, E2, E1, followed by a termination codon. The identification of the NH_2 terminus and the COOH terminus of each protein is discussed below. The deduced amino acid sequence is in precise agreement with the NH_2-terminal sequences found by automated sequence analysis of E1 and E2 (6), of PE2 (the precursor to E3 and E2) (J.R. Bell and M.W. Hunkapiller, personal communication), and of the hydrophobic "roots" of E1 and E2 derived by treatment of intact virions

Fig. 1. Nucleotide sequence of Sindbis 26S RNA and amino acid sequence of the encoded proteins. Nucleotides are numbered from the 3' end. Amino acids are numbered from the NH_2 terminus of each protein (C, E3, E2, 6K, E1). The beginning of each protein is labeled and the NH_2 terminus of the polypeptide segments from E2 and E1 isolated from spikeless particles after chymotrypsin treatment are marked with a triangle. Carbohydrate attachment sites are denoted by an asterisk. The single letter amino acid code is used: A = ala, C = cys, D = asp, E = glu, F = phe, G = gly, H = his, I = ile, K = lys, L = leu, M = met, N = asn, P = pro, Q = gln, R = arg, S = ser, T = thy, V = val, W = trp, Y = tyr.



with α-chymotrypsin (C.M. Rice and M.W. Hunkapiller, unpublished). In addition, sequences of many of the tryptic peptides from the Sindbis (strain Sa-AR-86) capsid protein have been recently published (7); most of this sequence data are consistent with our deduced sequence for the capsid protein.

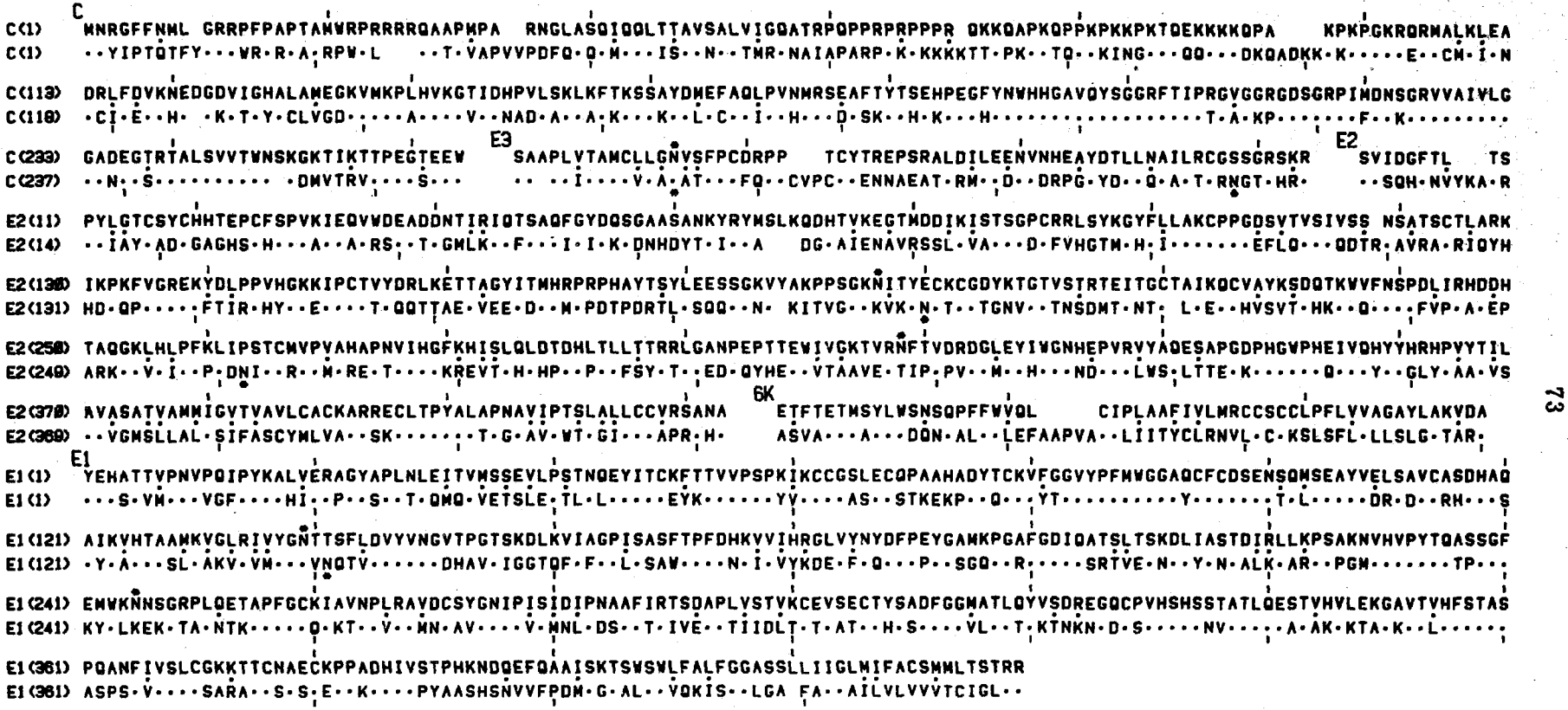
The experimental amino acid compositions for E1, E2, and C (8), are in excellent agreement with the compositions deduced from the nucleotide sequence and furnish additional support for the deduced protein sequences. In addition, our sequence data adjacent to the poly(A) tail is in complete agreement with the sequence obtained for the same region using a chain termination method (9).

The method we have used requires only a few micrograms of purified RNA, and because molecular cloning is not involved, there is no chance of selecting a minor variant in the population. We are currently using this method to locate sequence changes in Sindbis ts mutants.

The coding regions of 26S RNA of Semliki Forest virus (SFV) have been sequenced recently (10, 11). The amino acid sequences of Sindbis structural proteins are compared to those of SFV in Fig. 2. Sequences have been aligned to maximize homology. The overall homology between the viruses is striking: 47% of the residues are identical, and another 12% of the residues represent conservative substitutions. It is also obvious that some areas of the proteins are more highly conserved than others (Fig. 2), as will be discussed below. The cysteine residues are in general conserved except in certain hydrophobic areas (Fig. 2).

Overall Structure of Sindbis 26S RNA. We have identified a Hae III fragment of the 26S cDNA which includes the NH_2 terminus of the capsid protein and extends from the AUG codon 175 nucleotides towards the 5' end of the RNA. Assuming that this fragment is derived from the 26S RNA, this brings the overall length of the Sindbis 26S RNA to at least ~4230 nucleotides, not including the poly(A) tail, which is in good agreement with previous estimates of its size (2). A 5' untranslated region of

Fig. 2. Comparison of the amino acid sequence of the Sindbis structural proteins from Fig. 1 (upper sequence) and that of the structural proteins of SFV (from refs. 10, 11) (lower sequence). A dot in the SFV sequence means the amino acid is identical to that above in the Sindbis sequence. A dot between the amino acids of the two sequences denotes a conservative substitution (R=K, S=T, D=E, Q=N, V=L=I=M, A=G, A=V, Y=F). Possible carbohydrate attachment sites are marked with an asterisk. An attempt was made to keep the number of gaps introduced to a minimum. Nucleotide sequence homology was used to position gaps in some areas where amino acid sequence homology is low. The single letter amino acid code is used as in Figure 1.


C
E3
E2
6K
E1

175 nucleotides is fairly long in comparison with most known mRNAs; however, the mRNA for VP1 of SV40 has 240 nucleotides 5' to the initiation codon (12). The 3' untranslated sequence is 318 nucleotides in length in Sindbis compared to 264 nucleotides in SFV (11). It is of note that despite numerous deletions in the coding regions relative to one another (Fig. 2), the lengths of the two virus 26S RNAs are remarkably similar. From the AUG codon to the beginning of the poly(A) tail is 4023 nucleotides in SFV, and 4053 nucleotides in Sindbis.

We have examined the codon usage in Sindbis 26S RNA, and for most of the amino acids the distribution among the possible codons is clearly non-random (data not shown). The base composition or nearest neighbor frequency cannot account for the codon frequency found. In addition, in spite of the extensive nucleotide sequence divergence between Sindbis and SFV, the two viruses show similar codon usage (data not shown). (As an example of this divergence, regions in which the amino acid sequence is totally conserved between the two viruses show an average of about 0.8 base changes per codon.) We do not find a low frequency of the CG doublet as has been found for some DNA viruses such as SV40 (13), or RNA viruses such as fowl plague virus (14) which have mammalian or avian hosts. Sindbis and SFV have a wide host range which includes insects as well as birds and mammals. The codon preferences found in these two viruses may be due in part to an adaptation for optimal translational efficiency in this wide range of hosts.

The Capsid Protein. The capsid protein of Sindbis is 264 amino acids in length. Inspection of the sequence reveals a striking clustering of lysine, arginine, and proline in the NH_2-terminal half. Seventy percent of these three amino acids, as well as 85% of the glutamine and 60% of the alanine, are found within the first 120 residues. This basic region of the capsid protein is probably important in interacting with the virion RNA. Although SFV capsid protein has a similar basic region near the NH_2 terminus, there is relatively little sequence homology between the two proteins in this region

(Fig. 2). The COOH-terminal end (residues 166-264), however, shows remarkable sequence homology between Sindbis and SFV (Fig. 2) with 76% of the residues identical and another 6% conservative substitutions. This conservation suggests that this region could be important in protein-protein interactions such as those between capsid protein subunits to form the nucleocapsid, and those between the nucleocapsid and the COOH-terminal end(s) of the transmembrane viral glycoproteins.

Glycoprotein E3. Sindbis E3 is 64 amino acids in length. The first 19 residues are uncharged, highly conserved between Sindbis and SFV (Fig. 2), and may serve as the signal sequence for the insertion of PE2 into the endoplasmic reticulum during protein synthesis (15). This putative signal sequence is not cleaved from PE2 during protein synthesis; rather PE2 is cleaved to E2 and E3 during virus maturation (2, 15).

Sindbis E3 is known to be glycosylated (26) and the sequence contains a single glycosylation site of the type Asn-x-Ser/Thr (Asn_{14}). In addition, during amino acid sequence analysis of PE2 and E3, Asn_{14} is not recovered, as is characteristic of glycosylated asparagine residues, whereas other amino acids around this site are recovered (M.W. Hunkapiller, J.R. Bell and J. Mayne, personal communication). Thus it is virtually certain that Asn_{14} is glycosylated. The polysaccharide chain is known to be of the complex type (E.G. Strauss, personal communication). It is interesting to note that this glycosylated site occurs within the putative signal sequence for PE2, and is conserved between Sindbis and SFV (Fig. 2). The E3 polypeptide of SFV has in addition a second potential glycosylation site at Asn_{60} (11).

Glycoprotein E2. The E2 polypeptide of Sindbis is 423 amino acid residues long. There are two potential glycosylation sites. Burke and Keegstra (16) have shown that Sindbis E2 has two carbohydrate units, one of which is a simple oligosaccharide chain containing only mannose and N-acetylglucosamine and the other a complex type carbohydrate chain which contains in addition galactose, fucose, and sialic acid. They isolated glycopeptides following pronase digestion of E2 and found that the

glycopeptide of the complex type had the amino acid composition (Asn_1, $Ile_{0.8}$, $Thr_{0.7}$), while the simple type had the composition (Asn_1, $Phe_{0.6}$, $Thr_{0.1}$). From Figure 1 it is apparent that Asn_{196} carries a complex chain, and Asn_{318} has a simple chain. It is of interest that SFV E2 also carries two carbohydrate chains (11), one at Asn_{200}, the second at Asn_{262}.

Sindbis E2 has two long stretches of uncharged amino acids near its COOH terminus. The first of these is 28 amino acids long (residues 363 to 390) and begins near the NH_2 terminus of the hydrophobic "root" derived from E2 by treatment of the intact virion with chymotrypsin (unpublished) (Fig. 1). Since it is known that about 30 residues can be removed from the COOH-terminal end of PE2 by chymotrypsin treatment of intact microsomes (17), it is clear that this hydrophobic domain traverses the bilayer. E2 is known to have 5-6 residues of fatty acid covalently attached (18), all of which is found in the root (our unpublished data). There are three threonines and one serine in this first hydrophobic domain which could serve as attachment sites.

The second hydrophobic domain is 23 amino acids long (residues 396 to 418), and also appears to contain attached fatty acid (J.R. Bell, personal communication). Its orientation in the virion is unknown. The homology (87%) between the Sindbis and SFV proteins in this region is striking and contrasts with the low homology found in the first hydrophobic region. This homology together with the many conserved cysteine residues suggests that this segment may provide the specificity for the interaction between the viral nucleocapsid and the glycoproteins during budding.

55 Amino Acid Peptide (4.2 K or 6 K Protein). This peptide is made up of predominantly uncharged amino acids (91%) and probably serves in whole or in part as the signal sequence for E1. It is unknown if it has other functions as well. The corresponding Sindbis and SFV peptides have diverged widely (Fig. 2), implying that hydrophobicity is important for the function of this peptide, but the exact sequence is not.

Glycoprotein E1. E1, the viral hemagglutinin (19), is the largest of the Sindbis glycoproteins (439 amino acids). Two potential glycosylation sites (Asn_{139}, Asn_{245}) are present, and both appear to be glycosylated. Firstly, E1 of Sindbis grown in primary chicken cells contains both simple and complex polysaccharide chains, although E1 from Sindbis grown in a continuous hamster cell line contains only complex chains (20, 21). Secondly, the carbohydrate composition of E1 suggests two chains are present (21). Thirdly, the fact that Sindbis E1 migrates more slowly than E2 in SDS containing gels, whereas SFV E1 (which contains a single glycosylation site [11]) and E2 comigrate, suggests that Sindbis E1 has two chains as does E2. We hypothesize that Asn_{139} has a complex chain and Asn_{245} has a simple chain in chick grown virus, and both have complex chains in hamster grown virus. The single glycosylation site of SFV E1, Asn_{141} (11) is shifted by two residues from the first Sindbis site, a shift that required mutations in two separate codons (Fig. 2).

Sindbis E1 has two long sequences of uncharged amino acids. The first is 17 in length and begins at Val_{80}. In the region of E1 from Asp_{75} to Glu_{109}, which includes this uncharged segment, there are only 7 differences between Sindbis and SFV, and 4 of these are highly conservative changes (see Fig. 2). This uncharged region is not present in Sindbis spikeless particles (unpublished), suggesting that it is not imbedded in the viral membrane. It may play a role in mediating virus fusion with intracellular membranes during penetration (11).

The second region is located at the COOH terminus of E1 and begins seven residues from the NH_2 terminus of the root of E1 (unpublished) (Fig. 1). A sequence of 33 uncharged amino acids is followed by two arginine residues suggesting that E1 spans the bilayer with only two (or at most a few) residues exposed on the cytoplasmic side. This uncharged domain contains six serine and three threonine residues which are potential attachment sites for the 1-2 palmitic acid molecules located in this region of E1 (18; unpublished data). Sindbis and SFV show little

homology in this region, but both sequences are highly hydrophobic (Fig. 2).

Sites of Cleavage. The NH_2 terminus of the Sindbis capsid protein is Met-Asn-Arg- (7, 15; J.R. Bell, personal communication), thus identifying the start of this protein. The COOH terminus is established by the (capsid) tryptic peptide Thr-Thr-Pro-Glu-Gly-Thr-Glu-Glu-Trp sequenced by Boege, et al. (7), found at residues 256 to 264 (Fig.1). The NH_2-terminal sequence of PE2 (and E3), beginning Ser-Ala-Ala-Pro- (J.R. Bell, personal communication), follows directly (Fig. 1). Thus, tryptophan is the COOH-terminal amino acid of the capsid protein, and the protease responsible for the capsid-PE2 cleavage has a specificity similar to chymotrypsin. It has been suggested that this proteolytic activity resides in the capsid protein itself (22, 23, 24). The highly conserved region around the cleavage site (Fig. 2) may be important for a site-specific, viral encoded cleavage. This cleavage occurs during translation and apparently must take place if the signal sequence for PE2 is to function (25).

The cleavage site between E3 and E2 has yet to be precisely determined. The amino terminus of E2 is Ser-Val-Ile- (6) (Fig. 1). Since E3 is radiolabeled with lysine (J. Mayne, personal communication), the COOH terminus is either Lys or Lys-Arg. In SFV the COOH-terminal Arg is apparently removed (11). The cleavage of PE2 to form E3 and E2, in either case, involves a trypsin-like specificity. The origin and localization of the protease responsible for this is at present unknown. It occurs relatively late, approximately 20 minutes after synthesis of PE2 and is required for virus maturation (2). After cleavage Sindbis E3 is found in the culture medium (26), while that of SFV remains associated with the mature virion (27). Thus the cleavage may occur in the plasma membrane, outside the cell, or it may occur internally, perhaps in the lumen of Golgi vesicles shortly before or concomitant with their fusion with the plasma membrane. The latter alternative is suggested by analogy to the processing of such proteins as pro-albumin and pro-parathyroid hormone (11), which are cleaved after double basic amino acids in the Golgi shortly before secretion of the

mature protein (28). The failure of several groups to detect PE2 in the plasma membrane lends further support to this idea.

The COOH terminus of E2 is (Ala, Asn)-Ala, as determined with carboxypeptidase Y (T. Crowley, personal communication). The beginning of E1 is Tyr-Glu-His- (6) (Fig. 1). There is a sequence of 55 residues between the end of E2 and the beginning of E1 (Fig. 1). This is probably the peptide isolated by Welch and Sefton (26) from Sindbis virus infected cells, which they called a 4.2 K polypeptide. They have located the leucine and methionine residues in the first 20 residues of the 4.2 K polypeptide by microsequence analysis, and found methionine at position 7 and leucine at position 10 (B. Sefton, personal communication), in agreement with the sequence shown in Fig. 1.

The cleavages at the end of PE2 and the beginning of E1 both occur after alanine residues (for SFV as well as Sindbis, Fig. 2). It is thus tempting to propose that signalase is responsible for both cleavages (the properties of signalase have been recently reviewed in Ref. 29). If so, and if the signalase activity is restricted to the lumen of the rough endoplasmic reticulum, it would predict that the COOH-terminal regions of PE2 span the membrane twice, at least transiently. This could occur if the COOH terminus of PE2 and the 55 amino acid polypeptide form a set of stop transfer sequences and internal signal sequences that lead to multiple crossings of the membrane of the endoplasmic reticulum. After cleavage, the 55 amino acid peptide could be membrane-associated or could be released into the lumen of the endoplasmic reticulum. The COOH terminus of PE2, after cleavage, could remain transmembranous, or it could fold in such a way as to become partly or completely cytoplasmic. This domain contains proline and asparagine which would be unusual in an intramembranous region, but it also appears to contain covalently attached lipid (J.R. Bell, personal communication), suggesting membrane association. Further work will be required to resolve these questions. It is noteworthy, however, that the two

host proteases thought to be involved in processing of the Sindbis structural polypeptides are located in the lumen of the endoplasmic reticulum and in the Golgi apparatus, whereas a virus encoded protease appears to function in the cytosol. Virus specific proteases which cleave capsid precursors located in the cytosol have also been reported for other viruses (30), including picornaviruses (31), RNA tumor viruses (32), and probably adenovirus (33). We suggest that it is a general rule that virus encoded proteases are utilized for the processing in the cytosol of virus polypeptides.

ACKNOWLEDGEMENTS

We would like to thank E. Lenches for preparing the chick fibroblasts, T.E. Crowley for help with the 26S RNA isolation, H.V. Huang, T. Hunkapiller, and M. and R. Douglas for writing the computer programs and preparation of the figures, J. R. Bell, M.W. Hunkapiller, and E.G. Strauss for helpful discussions, and L. Hood and T. Maniatis for their generosity in lending us supplies and equipment. We are especially grateful to H. Garoff, W.J. Welch, and B.M. Sefton for sending us preprints of their work prior to publication. This work was supported by Grants PCM 77-26728 from the National Science Foundation, by Grants GM 06965 and AI 10792 from NIH, and by Biomedical Research Support Grant 507 RR 07003 from NIH. CMR was supported by Training Grant GM 00086 from NIH.

1. Strauss, E.G. (1978) J. Virol. **28**, 466-474.

2. Strauss, J.H. and Strauss, E.G. (1977) in The Molecular Biology of Animal Viruses, ed. Nayak, D. (Dekker, New York), Vol. 1, pp. 111-166.

3. Strauss, E.G. and Strauss, J.H. (1980) in The Togavirusus, ed. Schlesinger, R.W. (Academic Press, New York), pp. 393-426.

4. Maxam, A.M. and Gilbert, W. (1980) in Methods of Enzymology, eds. Grossman, L. and Moldave, K. (Academic Press, New York), Vol. 65, pp. 499-560.

5. Smith, D.R. and Calvo, J.M. (1980) Nucleic Acids Res. **8**, 2255-2274.

6. Bell, J.R., Hunkapiller, M.W., Hood, L.E. and Strauss, J.H. (1978) Proc. Natl. Acad. Sci. USA **75**, 2722-2726.

7. Boege, U., Wengler, G., Wengler, G. and Wittman-Liebold, B. (1980) Virology **103**, 178-190.

8. Bell, J.R., Strauss, E.G. and Strauss, J.H. (1979) Virology **97**, 287-294.

9. Ou, J., Strauss, E.G. and Strauss, J.H. (1981) Virology, **109**, 281-289.

10. Garoff, H., Frischauf, A-M., Simons, K., Lehrach, H. and Delius, H. (1980) Proc. Natl. Acad. Sci. USA **77**, 6376-6380.

11. Garoff, H., Frischauf, A-M., Simons, K., Lehrach, H. and Delius, H. (1980) Nature (London), **288**, 236-241.

12. Ghosh, P.K., Reddy, V.B., Swinscoe, J., Choudary, P.V., Lebowitz, P. and Weissman, S.M. (1978) J. Biol. Chem. **253**, 3643-3647.

13. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J. Volckaert, G. and Ysebaert, M. (1978) Nature (London) **273**, 113-120.

14. Porter, A.G., Barber, C., Carey, N.H., Hallewell, R.A., Threlfall, G. and Emtage, J.S. (1979) Nature (London) **282**, 471-477.

15. Bonatti, S. and Blobel, G. (1979) J. Biol. Chem. **254**, 12261-12264.

16. Burke, D. and Keegstra, K. (1979) J. Virol. **29**, 546-554.

17. Wirth, D.F., Katz, F., Small, B. and Lodish, H.F. (1977) Cell **10**, 253-263.

18. Schmidt, M.F.G., Bracha, M. and Schlesinger, M.J. (1979) Proc. Natl. Acad. Sci. USA **76**, 1687-1691.

19. Dalrymple, J.M., Schlesinger, S. and Russell, P.K. (1976) Virology **69**, 93-103.

20. Sefton, B.M. and Keegstra, K. (1974) J. Virol. **14**, 522-530.

21. Burke, D.J. and Keegstra, K. (1976) J. Virol. **20**, 676-684.

22. Simmons, D.T. and Strauss, J.H. (1974) J. Mol. Biol. **86**, 397-409.

23. Scupham, R.K., Jones, K.F., Sagik, B.P. and Bose, H.R., Jr. (1977) J. Virol. **22**, 568-571.

24. Aliperti, G. and Schlesinger, M.J. (1978) Virology **90**, 366-369.

25. Wirth, D.F., Lodish, H.F. and Robbins, P.W. (1979) J. Cell Biol. **81**, 154-162.

26. Welch, W.J. and Sefton, B.M. (1979) J. Virol. **29**, 1186-1195.

27. Garoff, H., Simons, K. and Renkonen, O. (1974) Virology **61**, 493-504.

28. Dean, R.T. and Judah, J.D. (1980) in Comprehensive Biochemistry, eds. Florkin, M. and Neuburger, A. (Elsevier, Amsterdam), Vol. 19B, pt. 1, pp. 233-298.

29. Blobel, G., Walter, P., Chang, C., Goldman, B.M., Erickson, A.H. and Lingappa, V.R. (1979), in Symposia of the Society for Experimental Biology, No. 33, pp. 9-36.

30. Butterworth, B.E. (1977) in Current Topics in Microbiology and Immunology, ed. Arber, W. (Springer-Verlag, Berlin), Vol. 76, pp. 1-41.

31. Palmenberg, A.C., Pallansch, M.A. and Rueckert, R.R. (1979) J. Virol. **32**, 770-778.

32. von der Helm, K. (1977) Proc. Natl. Acad. Sci. USA **74**, 911-915.

33. Bhatti, A.R. and Weber, J. (1979) Virology **96**, 478-485.

CHAPTER 4

Isolation and characterization of the hydrophobic COOH-terminal domains of the Sindbis virion glycoproteins

CHARLES M. RICE, MICHAEL W. HUNKAPILLER,

ELLEN G. STRAUSS and JAMES H. STRAUSS

Prepared for submission to the

Journal of Molecular Biology

ABSTRACT

Digestion of intact Sindbis virions with α-chymotrypsin produced a single membrane-associated peptide derived from each of the two virion glycoproteins (referred to as RE1 and RE2, or roots derived from E1 and E2, respectively). Amino acid composition data and NH_2-terminal sequence analysis established their location at the extreme COOH-terminal end of each glycoprotein. RE1 and RE2 are rich in hydrophobic amino acids and insoluble in aqueous solutions in the absence of detergents, and show differential solubility in organic solvent systems designed for the extraction of lipids. Essentially all of the covalently attached palmitic acid associated with E1 and E2 was found to be clustered in their hydrophobic, membrane-associated roots. Beginning 6-7 residues from their NH_2 termini RE1 and RE2 contain uninterrupted sequences of hydrophobic amino acids (33 and 26 amino acids, respectively), similar in terms of amino acid composition and length to the transmembrane anchors found in other bitopic integral membrane proteins. Following these uncharged, intramembrane segments there are clusters of predominantly basic amino acids. By structural analogy to known transmembrane proteins, E1 probably spans the bilayer but contains only a few residues exposed on the inner face of the virion envelope. In contrast, E2 and PE2 (the precursor to E2) have been shown to span the bilayer completely, and these proteins contain an additional 33 COOH-terminal residues which could be either exposed on the cytoplasmic face of the lipid bilayer or which could loop back into the membrane. This region at the extreme COOH-terminal end of E2 contains a second uncharged domain (23 amino acids in length), is highly conserved between Sindbis virus and closely related Semliki Forest virus, and contains several uncharged residues not typically found in transmembrane polypeptide segments. Although its orientation is unknown, this region may be involved in the highly specific interaction of the transmembrane glycoproteins in the plasma membrane with the cytoplasmic nucleocapsid during budding.

INTRODUCTION

Alphaviruses such as Sindbis virus and the closely related Semliki Forest virus (SFV) are simple enveloped viruses useful for the study of glycoprotein biosynthesis. In the case of Sindbis virus, the mature virion contains two glycoproteins (E1 and E2) anchored in a lipid bilayer derived from the host cell, surrounding an icosahedral nucleocapsid assembled in the cytoplasm (Strauss and Strauss, 1977). The majority of the glycoprotein mass is external to the bilayer and can be removed by protease treatment (Compans, 1971; Garoff and Söderlund, 1978). All three structural proteins (E1, E2, and the capsid protein) are encoded by a single 26S subgenomic mRNA with a single initiation site for protein synthesis (Cancedda et al., 1975; Rice and Strauss, 1981). The genes are translated on membrane-bound polyribosomes (Wirth et al., 1977) in the order 5'-C-E3-E2-6K-E1-3' (Garoff et al., 1980a,b; Schlesinger and Kääriäinen, 1980; Rice and Strauss, 1981), as a continuous polypeptide which is processed by proteolytic cleavage (E3 and the 6K protein are not found in mature Sindbis virions). The nascent glycoprotein precursors are cotranslationally inserted into the lumen of the rough endoplasmic reticulum (Bonatti et al., 1979; Garoff et al., 1978), and core glycosylation with mannose-rich oligosaccharides takes place during or shortly after protein synthesis (Sefton, 1977). The glycoproteins then appear to follow the vectorial route of other secreted and plasma membrane proteins to the smooth endoplasmic reticulum and eventually to the plasma membrane (Palade, 1975; Erwin and Brown, 1980). During this time oligosaccharide chains are trimmed, and complex chains are produced by the addition of galactose, fucose, and sialic acid. In addition, both glycoproteins contain covalently attached fatty acids (Schmidt et al., 1979), which are added 10-20 minutes after the completion of protein synthesis (Schmidt and Schlesinger, 1980). PE2 (the precursor to E3 and E2) and E2 have been shown to be a transmembrane proteins since they can be shortened by about 30 amino acids by proteolysis of microsomes from infected cells (Wirth et al., 1977; Ziemiecki

et al., 1980) or heterologous cell-free systems (Garoff et al., 1978; Bonatti et al., 1979). The final events during maturation involve the specific interaction of the nucleocapsid with the cytoplasmic portion of the transmembrane glycoprotein(s) leading to virus budding through the host cell plasma membrane. The glycoproteins can be incorporated into mature virions about 20 minutes after their synthesis, and the cleavage of PE2 to E3 and E2, which also occurs at this time, is necessary for the final steps in virus maturation (Strauss and Strauss, 1977). This interaction between the virion glycoproteins and the nucleocapsid appears to be highly specific since host cell glycoproteins are rigorously excluded from mature virions (Strauss, 1978).

Sindbis virus, SFV (Garoff and Söderlund, 1978) and other simple enveloped viruses such as vesicular stomatitis virus (Katz et al., 1977; Rose et al., 1980), are easily tractable systems in which to study the topology of membrane glycoproteins. Protease treatment of intact virions yields the membrane-associated domains of the glycoproteins which can then be readily purified for further biochemical analysis. The transmembrane nature of these proteins, in some cases, can be verified directly by protease treatment of infected cell microsomes (Wirth et al., 1977; Katz et al., 1977). Primary sequence analysis of these membrane-associated peptides can then be used to establish boundaries for both the cytoplasmic and extracytoplasmic sides of the transmembrane domain. This approach would be difficult for most cellular transmembrane glycoproteins because of the difficulties involved in isolating small, protease-resistant transmembrane domains from such complex mixtures for subsequent sequence analysis. Thus, by studying viral glycoprotein transmembrane domains, it may be possible to establish general criteria which could be used for predicting transmembrane regions on the basis of primary structure for proteins more difficult to analyze directly.

In this study, we have produced spikeless particles (Uterman and Simons, 1974) by α-chymotrypsin digestion of intact Sindbis virions, and extensively characterized

the protease-resistant, membrane-associated glycoprotein fragments, called "roots." In previous studies, the entire glycoprotein sequences have been deduced by direct protein and cDNA sequence analysis (Bell et al., 1978; Rice and Strauss, 1981), thus allowing the precise localization of these hydrophobic segments in the glycoprotein sequence. These data show that each virion glycoprotein is anchored in the viral membrane by a transmembrane hydrophobic domain at or near the COOH terminus. E2 contains an additional sequence of 33 amino acids COOH-terminal to its transmembrane domain which is highly conserved between Sindbis virus and SFV and may play a role in the specific interaction of the spike glycoproteins with the nucleocapsid during budding.

MATERIALS AND METHODS

Virus growth and purification. Sindbis virus (HR strain) was grown in monolayers of primary chicken embryo fibroblasts as previously described (Pierce et al., 1974). All radiochemicals were purchased from Amersham Searle. Radiolabeled virus was prepared by the salt-reversal method and purified by sequential sedimentation velocity and isopycnic centrifugation (Pierce et al., 1974). Isopycnic gradients contained 200 μg/ml bovine serum albumin (BSA) as carrier. Labeling of virus with [9,10-^{3}H(n)] palmitic acid (500 μCi/mmole) was done essentially as described by Schmidt et al. (1979). Milligram quantities of virus were grown and purified by the method of Bell et al. (1979).

Preparation of spikeless particles. Radiolabeled virus (in 33% sucrose, 90% D_2O, 0.05 M Tris-Cl, pH 7.4, 0.2 M NaCl, 0.001 M EDTA, and 200 μg/ml BSA) was diluted with 2 volumes of 0.05 M Tris-Cl, pH 7.4, 0.2 M NaCl, 0.001 M EDTA (TNE buffer) containing 200 μg/ml BSA. α-Chymotrypsin (Worthington), from a fresh 20 mg/ml stock in TNE buffer, was added to a final concentration of 3 mg/ml. After incubation at 37°C for 90 minutes, the reaction was stopped by the addition of 1 mg

phenylmethyl sulfonyl fluoride (PMSF) (20 mg/ml in absolute ethanol) for each 10 mg of α-chymotrypsin. Samples were diluted with TNE buffer prior to the addition of PMSF such that the final ethanol concentration did not exceed 1%. After incubation for 10 minutes at room temperature, the samples were chilled on ice.

For amino acid composition and NH_2-terminal sequence analysis, larger quantities of spikeless particles were prepared from virus purified by the method of Bell et al. (1979). A small amount of ^{35}S-methionine labeled virus was included as a tracer. Pooled virus from the isopycnic gradient (1 mg/ml virus protein) was digested as described above except that the BSA carrier was omitted and digestion was for 70 minutes at 37°C.

Two methods were used for separation of the spikeless particles from the released proteolysis products. Samples for subsequent gel electrophoresis were pelleted through 3 ml of 15% sucrose in TNE buffer containing 200 µg/ml BSA and 100 µg/ml PMSF by centrifugation in a Spinco SW 50.1 rotor at 50K rpm for 3 hours at 4°C. Alternatively, spikeless particles were banded by isopycnic centrifugation on linear gradients containing 18-40% sucrose, 50% D_2O, 200 µg/ml BSA, and 20 µg/ml PMSF, in TNE buffer. After centrifugation in a Spinco SW 40 rotor at 32K rpm for 12 hours, the band was collected by puncturing the bottom of the tube.

Polyacrylamide gel electrophoresis. Polyacrylamide gel electrophoresis was performed using the discontinuous buffer system of Laemmli (1970) except that the buffer concentrations were halved. Slab gels contained either 10% acrylamide or an exponential gradient from 12-20% acrylamide (acrylamide:bisacrylamide, 30:0.8). After electrophoresis, analytical gels were treated for fluorography (Bonner and Laskey, 1974), and exposed at -70°C using pre-fogged Kodak X-Omat R film (Laskey and Mills, 1975). Preparative slab gels (for samples labeled with either ^{35}S- or ^{14}C- amino acids) were exposed at 4°C and the bands were excised and eluted into 0.02% sodium dodecyl sulfate (SDS), 2 mM dithiothreitol (DTT) and 20 µg/ml

PMSF. Cylindrical gels, 0.6 x 12 cm, contained either 10% acrylamide, as described above, or 20% acrylamide (Welch and Sefton, 1979). After electrophoresis, the gels were frozen and cut into 1 mm slices using a Mickle gel slicer. Preparative gel slices were shaken in 0.02% SDS, 2 mM DTT, 20 µg/ml PMSF, and aliquots were analyzed by liquid scintillation counting to localize the radioactive peaks. Gel slices for liquid scintillation counting were shaken for at least 48 hours in 10 ml of a toluene-based scintillation cocktail containing 5% NCS (Amersham) and 0.33% H_2O. Double-labeled samples were counted with appropriate standards and the channel overlap was corrected by a computer program.

Solubility in organic solvents. Organic solubility of gel purified samples was examined using two different systems. The first system involved extraction of the aqueous sample (containing 25 mg/ml carrier BSA) in 0.1 M NaCl, 0.05 M Tris-Cl, pH 7.4, and 1 mM EDTA (NTE) with 20 volumes of chloroform:methanol (2:1) according to the method of Folch et al. (1957). The second system used 10 volumes of acetone: ethanol (1:1) to one volume of the sample as described above, except that the NTE was omitted. After shaking for 1 hour, the pellet was removed by centrifugation at 3500 x g and washed with acetone:ethanol:H_2O (5:5:1). After flash evaporation or air drying, the samples were resuspended in 1% SDS and quantitated by liquid scintillation counting.

Peptide mapping. Protein samples, purified by polyacrylamide gel electrophoresis as described above, were desalted by centrifugation (Neal and Florini, 1973) using Sephadex G-25 (medium) equilibrated in 0.02% SDS. 100 µg of BSA was added and the samples were lyophilized. After performic acid oxidation (Hirs, 1967), the samples were lyophilized twice and resuspended in 100 µl of water with 400 µg BSA, and adjusted to 20% trichloroacetic acid. Following a 30-minute incubation on ice, three volumes of cold water were added and the trichloroacetic acid insoluble material was pelleted by centrifugation at 3500 x g for 6 minutes at 0°C. The pellets

were washed in 1% trichloroacetic acid and resuspended in 0.2 M NH_4HCO_3. Tryptic fragments were prepared by digestion at 37°C for 24 hours using two successive additions of 10% w/w trypsin [L(tosylamide-2-phenyl)-ethyl-chloromethyl-ketone treated, Worthington]. After removal of a portion of the sample, chymotrypsin and trypsin were added (10% w/w of each) and the digestion allowed to proceed for an additional 12 hours at 37°C. The samples were frozen, lyophilized, resuspended in 0.5 ml H_2O, and lyophilized a second time.

The peptides were resuspended in 100 µl 0.5 M $NaPO_4$:acetone (2:1), pH 1.8, centrifuged briefly, and separated on a DuPont 830 high pressure liquid chromatograph (HPLC) (McMillan et al., 1979). The column was equilibrated in 0.1 M $NaPO_4$, pH 2.1, and the peptides were eluted with an exponential gradient of acetone to 90% (approximated by the curve $y = x^3$) usually at 49°C. A 25 cm Zorbax-CN column was used for tryptic digests, whereas a 25 cm Zorbax-C18 column was used for separation of peptides after double digestion with both trypsin and chymotrypsin. Samples to be compared were mixed and co-chromatographed.

Amino acid composition. Nanomolar amounts of RE1 and RE2 were prepared by preparative polyacrylamide gel electrophoresis as described above. Samples were desalted twice by centrifugation (Neal and Florini, 1973) over Sephadex G-25 equilibrated in 0.02% SDS and dialyzed extensively against 0.02% SDS. After lyophilization, the samples were hydrolyzed for 24 hours with 6 N HCl at 110°C. Automated amino acid analyses of protein hydrolysates were performed on a Durrum D-500 amino acid analyzer. Cysteine and tryptophan were not determined.

An alternative method for determining the composition of RE1 and RE2 was based on incorporation of radiolabeled amino acids. Separate preparations of spikeless particles each labeled with a different amino acid were pelleted and analyzed on 20% polyacrylamide gels. The peaks corresponding to C, RE1, and RE2 were quantitated as described above. Assuming that these components were in equimolar amounts

(Schlesinger et al., 1972; Garoff et al., 1974), the number of amino acid residues in RE1 and RE2 were calculated using the amino acid composition and molecular weight data for the capsid protein (Bell et al., 1979; Rice and Strauss, 1981). This assumption has been verified for intact virions using several different amino acid labels (E. G. Strauss, unpublished data).

NH_2-terminal automated sequence analysis. Samples were prepared as described above for amino acid analysis. Approximately 2 nmoles of purified protein in 25% acetic acid was loaded onto an extensively modified Beckman Instruments spinning cup sequenator (Hunkapiller and Hood, 1978, 1980). Polybrene was used as a carrier and the instrument was run under a Quadrol protein program with double cleavage. Amino acid phenylthiohydantoins were separated and quantitated by reverse-phase HPLC on DuPont Zorbax ODS (Johnson et al., 1979).

RESULTS

Proteolysis of Sindbis virions. In preliminary experiments a variety of proteases were tested (including bromelain, α-chymotrypsin, papain, pronase, thermolysin, and trypsin) for their ability to digest the virion glycoproteins, and produce spikeless particles of lighter buoyant density. The results showed that Sindbis virion glycoproteins were highly resistant to proteolysis (data not shown). A ten-fold excess of α-chymotrypsin by weight was required to efficiently digest the virion glycoproteins (see below). α-Chymotrypsin was chosen for several reasons. The enzyme preferentially cleaves COOH-terminal to hydrophobic amino acid residues and would be expected to extensively degrade hydrophobic peptides released from virions thus preventing their nonspecific association with spikeless particles during further purification. Although not as specific as trypsin, the limit digestion of intact virions with α-chymotrypsin would probably produce membrane-protected peptides with unique or at least preferential NH_2 termini. In addition, the enzyme was easily inactivated by

the addition of PMSF. Figure 1A compares the isopycnic gradient profiles of ^{35}S-methionine and ^{3}H-glucosamine labeled Sindbis virus digested with α-chymotrypsin. 43% of the methionine label, but less than 1% of the glucosamine label remained associated with the spikeless particles. Further analysis by acrylamide gel electrophoresis (Fig. 1b) showed that the residual glycosamine label probably resulted from metabolic recycling into amino acids (see the capsid protein peak, Fig. 1B), from glycolipids, and from an unidentified component barely entering the gel. These results demonstrated that the glycoproteins had been efficiently and uniformly digested and the spikeless particles cleanly separated from the released proteolysis products.

Methionine labeled spikeless particles contained the capsid protein as well as three proteolysis products, C-1, RE1, and RE2 (see Figs. 1B and 2). The largest of these, C-1, was produced in significant amounts only after prolonged proteolysis which resulted in the loss of the capsid protein (data not shown), and has been shown to be a capsid protein degradation product (see below). Optimal digestion conditions for the production of RE1 and RE2 were determined empirically such that degradation of the capsid protein (appearance of C-1) was minimized. Short incubations with high α-chymotrypsin concentrations were preferable to longer incubations with reduced enzyme. In addition, virus isolated by the salt-reversal method (Pierce et al., 1974) was significantly less susceptible to leakage resulting in capsid protein degradation than was the virus isolated by the large scale preparative method of Bell et al. (1979). The basis for this difference is at present unknown, but may result from the exposure of the virus to fluorocarbon during the large scale purification (Bell et al., 1979).

The origin of these protease-resistant fragments in spikeless particles was determined by comparing their tryptic and α-chymotryptic peptides to those of the virion structural proteins. HPLC separation of the peptides was chosen because it allows the simultaneous comparison of two appropriately labeled samples. The elution profiles are extremely reproducible (usually to within 1 fraction out of 180)

Fig. 1. Preparation of Sindbis virus spikeless particles. Spikeless particles were prepared by α-chymotrypsin digestion of Sindbis virus labeled with ^{35}S-methionine (- - - - -) or ^{3}H-glucosamine (———) as described in the Materials and Methods. Panel A shows the isopycnic gradient profile of the proteolytic digestion products (the gradient was collected from the bottom). Portions of the spikeless particle peaks from parallel gradients were pooled and run on a 20% acrylamide gel as described in the Materials and Methods (panel B). Electrophoresis was from left to right.

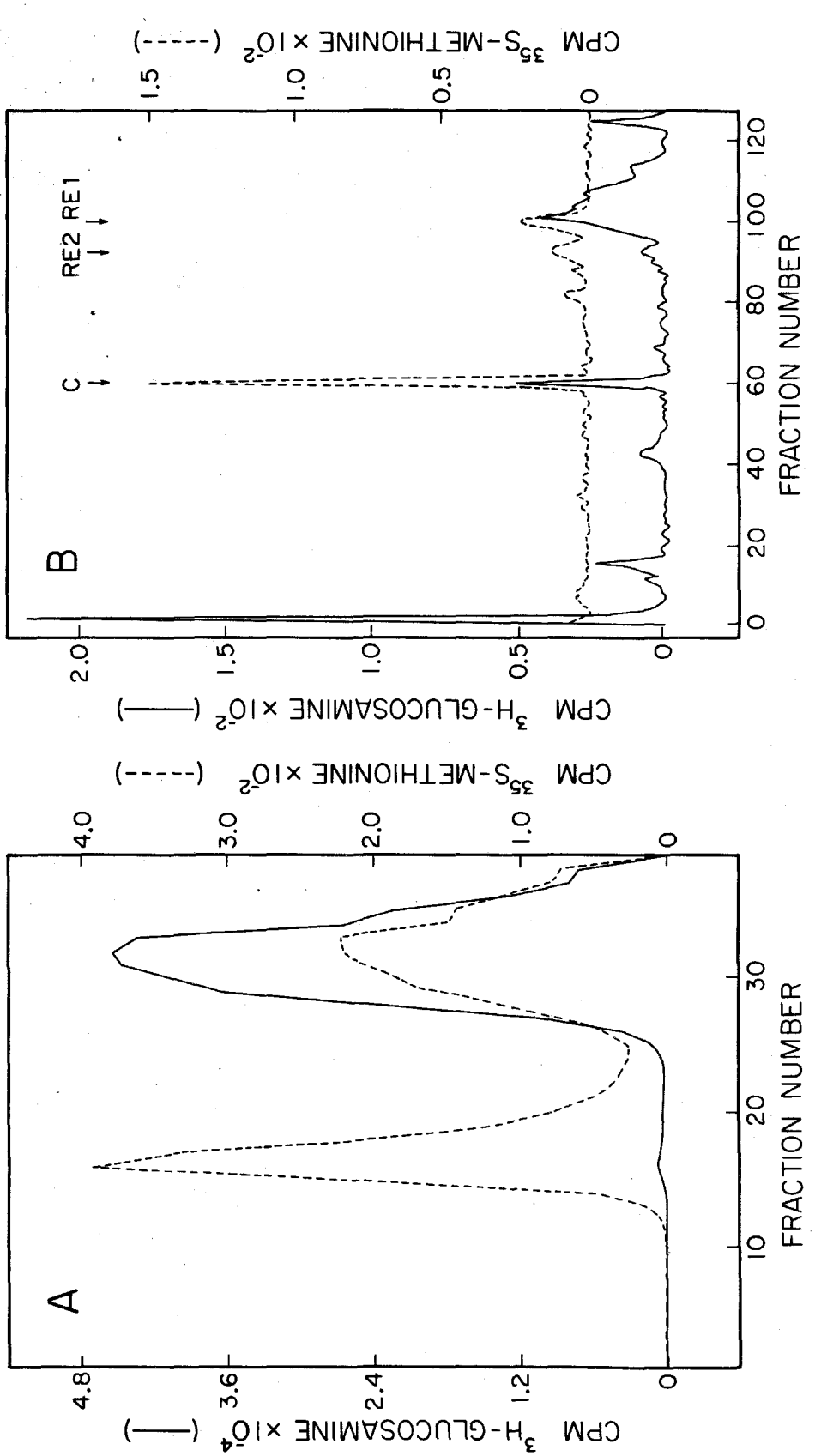
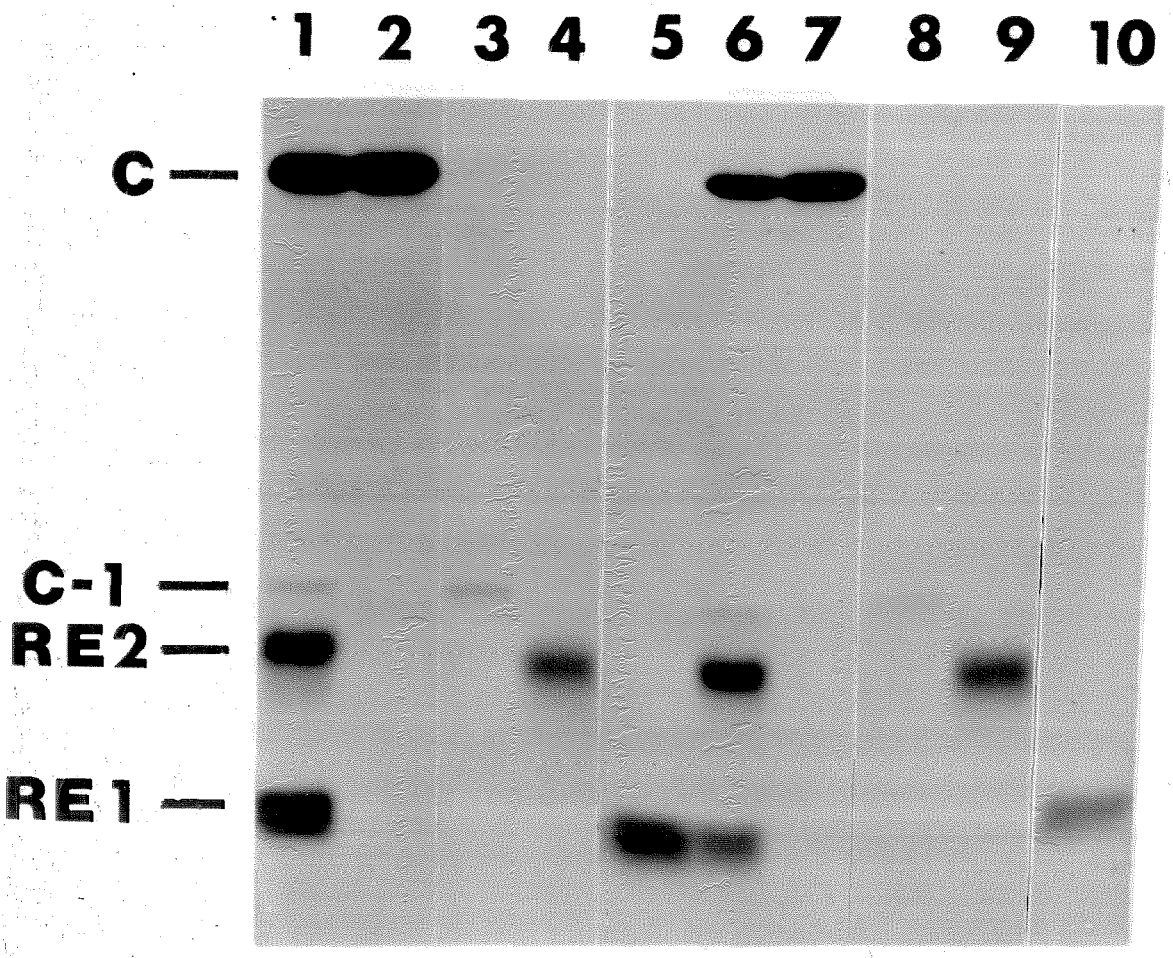

A
B
C
RE2
RE1
FRACTION NUMBER
CPM ^{3}H-GLUCOSAMINE ×10^{-4} (——)
CPM ^{35}S-METHIONINE ×10^{-2} (- - - -)
CPM ^{3}H-GLUCOSAMINE ×10^{-2} (——)
CPM ^{35}S-METHIONINE ×10^{-2} (- - - -)

Fig. 2. Gel purified proteins of Sindbis virus spikeless particles. ^{35}S-methionine (lanes 1-5) and ^{14}C-leucine (lanes 6-10) spikeless particles were prepared as described in the Materials and Methods. The polypeptide components were isolated by preparative electrophoresis on 12-20% acrylamide gradient gels. After elution from the gel slices, samples of the capsid protein (lanes 2 and 7), C-1 (lanes 3 and 8), RE2 (lanes 4 and 9), and RE1 (lanes 5 and 10) were rerun on a 12-20% acrylamide gradient gel. The original spikeless particle preparations are shown in lanes 1 and 6.


1 2 3 4 5 6 7 8 9 10
C
C-1
RE2
RE1

and can be analyzed for both qualitative and quantitative data. A preliminary experiment showed that RE1 and RE2 were tenaciously bound to the HPLC column and were not eluted with 90% acetone (data not shown). This result was not unexpected since the column separation is based upon hydrophobic interactions, and these membrane-associated peptides were extremely hydrophobic (see below). Analysis of tryptic fragments of RE1 and RE2 by gel electrophoresis and two-dimensional thin layer separation showed that only a small number of large, presumably hydrophobic, leucine- and methionine-containing peptides were produced (data not shown). Thus, to insure the cleavage of long hydrophobic peptides into smaller fragments which could be eluted from the HPLC column, the proteins were usually digested with a combination of trypsin and α-chymotrypsin. Figure 3 shows several representative peptide maps for leucine- and methionine-containing peptides. These results show that C-1 is derived from the capsid protein whereas RE1 and RE2 are derived from E1 and E2, respectively.

Amino acid compositions. The amino acid compositions of RE1 and RE2 were determined by two independent methods and are presented in Table 1. For comparison, the compositions based on the complete sequences of RE1 and RE2 are included (see below; Rice and Strauss, 1981). The compositions determined by automated amino acid analysis of RE1 and RE2 isolated from large-scale virus preparations (Bell et al., 1979) contained significant contamination. The difficulties in obtaining clean preparations of spikeless particles from these virus preparations without extensive degradation of the capsid protein have been mentioned. Capsid degradation products or autocatalytic degradation products from the α-chymotrypsin copurifying with spikeless particles and comigrating with RE1 and RE2 during gel electrophoresis could explain these results. The capsid protein prepared in parallel with RE1 and RE2 gave a composition in good agreement with previously published data (Bell et al., 1979). The second method was based on the known molecular weight and composition

Fig. 3. HPLC peptide maps of polypeptides from spikeless particles and intact virions. Gel purified spikeless particle polypeptides (C-1, RE1, and RE2) were prepared as described in Fig. 2. Tritium-labeled virion structural proteins (C, E1, and E2) were separated on cylindrical 10% acrylamide gels and eluted. After desalting, and performic acid oxidation, the samples were digested with either trypsin (T) (panel C) or a combination of both trypsin and α-chymotrypsin (CT+T) (panels A,B,D,E, and F). The peptides were resolved by high pressure liquid chromatography on either a Zorbax- C18 column (panels A,B,D,E, and F) or a Zorbax-CN column (panel C) as described in the Materials and Methods. Chromatography was at 49°C, except in the case of panel A which was run at 23°C. Since they were digested and chromatographed under identical conditions, the samples in panels B,D,E, and F are comparable. (———) Labeled peptides derived from intact structural proteins. (-----) Labeled peptides derived from C-1, RE1 or RE2.

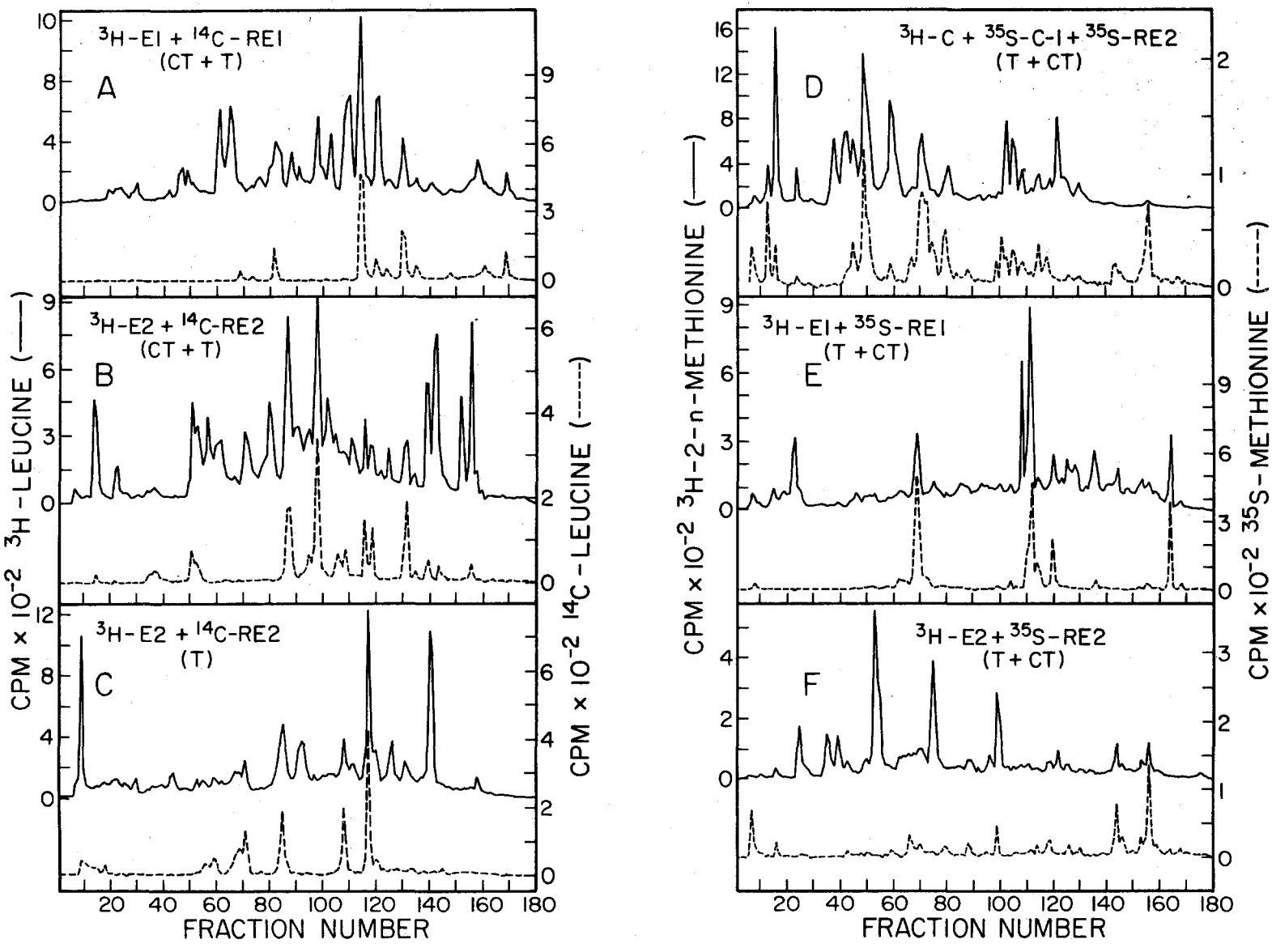

A
^{3}H-E1 + ^{14}C-RE1
(CT + T)
B
^{3}H-E2 + ^{14}C-RE2
(CT + T)
C
^{3}H-E2 + ^{14}C-RE2
(T)
D
^{3}H-C + ^{35}S-C-1 + ^{35}S-RE2
(T + CT)
E
^{3}H-E1 + ^{35}S-RE1
(T + CT)
F
^{3}H-E2 + ^{35}S-RE2
(T + CT)
CPM x 10^{-2} ^{3}H-LEUCINE (——)
CPM x 10^{-2} ^{14}C-LEUCINE (- - - -)
CPM x 10^{-2} ^{3}H-2-n-METHIONINE (——)
CPM x 10^{-2} ^{35}S-METHIONINE (- - - -)
FRACTION NUMBER

Table 1

Amino Acid Analysis of RE1 and RE2

	RE1			RE2		
Amino Acid	residues[a]	residues[b]	residues[c]	residues[a]	residues[b]	residues[c]
Asp	0	-	2.1	0	-	3.9
Asn	0	-		2	-	
Thr	3	3.4	6.2	5	5.4	5.5
Ser	7	5.9	n.d.[e]	3	2.6	3.8
Glu	0	-	3.5	1	-	4.1
Gln	1	-		0	-	
Pro	0	0[d]	1.2	4	3.8	5.9
Gly	3	-	n.d.[f]	1	0	n.d.[f]
Ala	5	-	9.2	13	-	15.0
Cys	1	-	n.d.	5	-	n.d.
Val	0	0[d]	1.3	8	7.7	9.2
Met	3	3.0	3.0	2	2.1	2.0
Ile	4	3.9	2.7	3	2.9	3.1
Leu	6	6.2	7.5	7	6.9	8.5
Tyr	0	0[d]	0.3	3	2.8	2.8
Phe	3	2.7	3.4	0	0[d]	0.7
His	0	0[d]	0.2	2	1.7	2.5
Lys	1	1.0	2.1	1	1.1	2.6
Arg	2	2.1	3.4	4	4.4	5.6
Trp	2	1.9	n.d.	0	0[d]	n.d.
Total	41			64		

[a]Amino acid composition deduced from the 26S RNA sequence (Rice and Strauss, 1981).

[b]Amino acid composition determined by incorporation of radiolabeled amino acids as described in the Material and Methods and the text.

[c]Results from automated amino acid analysis. n.d. = not determined.

[d]No discrete peak of radioactivity was found in the proper position for these amino acids.

Table 1 (continued)

[e]Not determined because of serine containing phospholipid contamination.

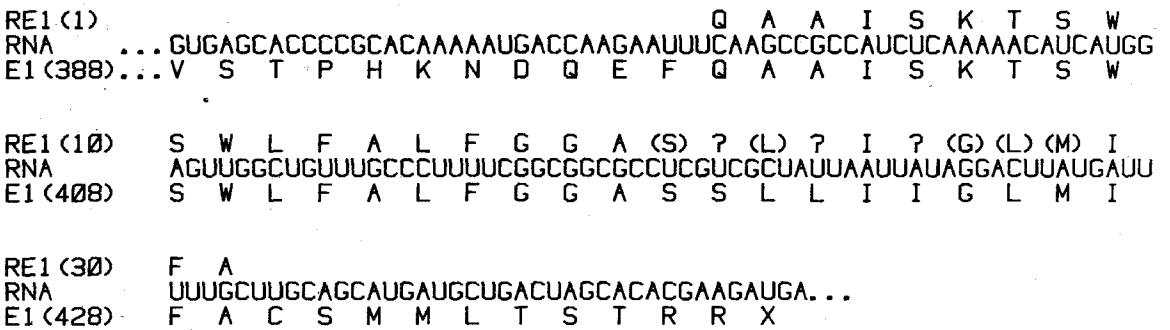
[f]Not determined because of residual glycine in the sample from preparative electrophoresis.

of the capsid protein, and a comparison of ratios of amino acid label in RE1, RE2, and C found in spikeless particles. Assuming an equimolar ratio of these components in the spikeless particle (and therefore equimolar amounts of E1, E2 and C), the number of residues of a given amino acid in RE1 or RE2 is simply the fraction of that labeled amino acid in the root compared to the capsid protein multiplied by the number of residues of the amino acid in the capsid protien. These compositions are in excellent agreement with the compositions proposed on the basis of protein and nucleic acid sequence data (see below; Rice and Strauss, 1981). This also verifies that E1, E2, and the capsid protein are present in very close to equimolar amounts in mature virions. This method was insensitive to unlabeled contaminants, and allowed the use of small quantities of virus prepared by the salt-reversal method (Pierce et al., 1974) from which clean preparations of spikeless particles could be produced. As expected, both RE1 and RE2 contain predominantly uncharged, hydrophobic amino acids. RE2 has more charged amino acids (5 basic and 1 acidic) which may explain its more limited solubility in organic solvents when compared to RE1 (see below). Both roots contain sufficient serine, threonine, or tyrosine to serve as attachment sites for fatty acid molecules (see below).

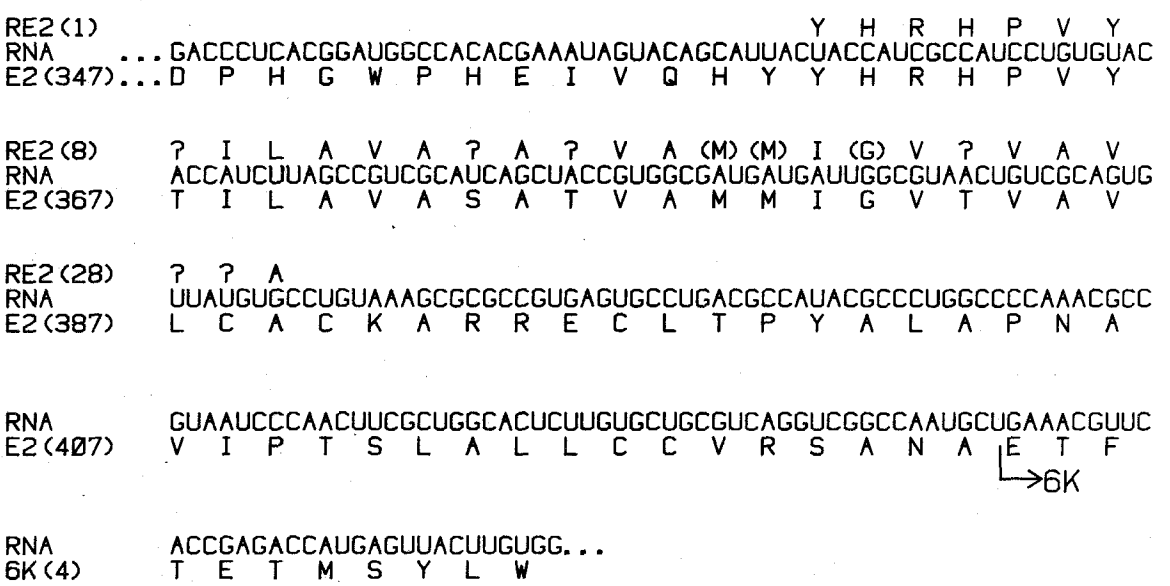
NH_2 terminal sequence analysis of RE1 and RE2. In order to determine the location of RE1 and RE2 in the parent glycoprotein molecules, two nanomoles of each root were subjected to sequential Edman degradation using an automated sequenator. These results are shown in Fig. 4. For purposes of comparison, the NH_2-terminal sequence data are aligned with portions of E2, the 6K protein, and E1 deduced from the sequence analysis of complementary DNA to the 26S RNA of Sindbis virus (Rice and Strauss, 1981). RE1 and RE2 begin 41 and 64 amino acids from the COOH-terminal ends of E1 and E2, respectively. The localization of the COOH termini of the glycoproteins and their roots is discussed below. The α-chymotryptic cleavages generating RE1 and RE2 both occur after bulky hydrophobic amino acids (phenylalanine and

Fig. 4. NH_2-terminal sequence analysis of RE1 and RE2. The NH_2-terminal amino acid sequences of RE1 and RE2 obtained by automated Edman degradation are compared to the polypeptide sequences deduced from the 26S RNA sequence (Rice and Strauss, 1981). The numbers indicate the position from the amino terminus of the protein. Tentative amino acid assignments are enclosed by parentheses, and unidentified residues are indicated by question marks. The single letter amino acid code is used: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; X, termination codon; Y, Tyr.

RE1


RE1(1) Q A A I S K T S W
RNA ...GUGAGCACCCCGCACAAAAAUGACCAAGAAUUUCAAGCCGCCAUCUCAAAAACAUCAUGG
E1(388)...V S T P H K N D Q E F Q A A I S K T S W
RE1(10) S W L F A L F G G A (S) ? (L) ? I ? (G) (L) (M) I
RNA AGUUGGCUGUUUGCCCUUUUCGGCGGCGCCUCGUCGCUAUUAAUUAUAGGACUUAUGAUU
E1(408) S W L F A L F G G A S S L L I I G L M I
RE1(30) F A
RNA UUUGCUUGCAGCAUGAUGCUGACUAGCACACGAAGAUGA...
E1(428) F A C S M M L T S T R R X

RE2


RE2(1) Y H R H P V Y
RNA ...GACCCUCACGGAUGGCCACACGAAAUAGUACAGCAUUACUACCAUCGCCAUCCUGUGUAC
E2(347)...D P H G W P H E I V Q H Y Y H R H P V Y
RE2(8) ? I L A V A ? A ? V A (M) (M) I (G) V ? V A V
RNA ACCAUCUUAGCCGUCGCAUCAGCUACCGUGGCGAUGAUGAUUGGCGUAACUGUCGCAGUG
E2(367) T I L A V A S A T V A M M I G V T V A V
RE2(28) ? ? A
RNA UUAUGUGCCUGUAAAGCGCGCCGUGAGUGCCUGACGCCAUACGCCCUGGCCCCAAACGCC
E2(387) L C A C K A R R E C L T P Y A L A P N A
RNA GUAAUCCCAACUUCGCUGGCACUCUUGUGCUGCGUCAGGUCGGCCAAUGCUGAAACGUUC
E2(407) V I P T S L A L L C C V R S A N A E T F
6K
RNA ACCGAGACCAUGAGUUACUUGUGG...
6K(4) T E T M S Y L W

tyrosine, respectively). RE1 contains a charged residue (lysine) at sixth position from the amino terminus followed by an uninterrupted stretch of 33 uncharged amino acids. RE2 contains a proline at position 5 followed by a sequence of 26 uncharged amino acids typical of transmembrane polypeptide segments (see Discussion). After this first hydrophobic segment, there is a cluster of charged residues (4 out of 5) followed by another stretch of 23 uncharged amino acids containing several prolines.

Determination of the COOH termini of RE1, RE2, E1, and E2. Given the COOH-terminal location of RE1 and RE2, it was of interest to determine and compare the COOH-termini of the glycoproteins and their roots. In the case of RE1, the presence of two arginines at the COOH terminus (see Fig. 4) can be inferred from the composition data presented in Table 1. Since a termination codon in the RNA sequence immediately follows these two arginines (Fig. 4), RE1 and E1 must have the same COOH terminus, -Leu-Arg-Arg-COOH. The localization of the COOH termini of RE2 and E2 was more difficult and follows from lines of evidence using several different approaches. The first approach utilized limited acid-catalyzed cleavage of E2. Other investigators have shown that aspartyl-proline bonds are particularly susceptible to acid-catalyzed cleavage (Piszkiewicz et al., 1970). The only such bond in E2 (Rice and Strauss, 1981; see Fig. 4), Asp_{347}-Pro_{348}, was efficiently cleaved by 70% formic acid (at 37°C for 24 hours) yielding a discrete COOH-terminal fragment (about 10,000 daltons) which was not labeled with phenylalanine (data not shown). Thus, the COOH terminus of E2 could not extend beyond the threonine shown in position 2 of the 6K protein (see below, Fig. 4). Microsequence analysis of the 6K protein (W. Welch and B. Sefton, personal communication), the gene translated immediately following E2, has revealed a methionine residue at position 7 and a leucine residue at position 10 (see Fig. 4). These data are consistent with a proteolytic cleavage after Ala_{423} generating the COOH terminus of E2 and the NH_2 terminus of the 6K protein. The COOH terminus of E2 when directly examined by digestion

with carboxypeptidase Y appears to be (Ala,Asn)-Ala-COOH (T. Crowley, personal communication). The COOH terminus of RE2 has been less clearly defined. The composition data shown in Table 1 show that RE2 contains 3 serine and 4 arginine residues. The third serine and fourth arginine in the RE2 sequence are the fourth and fifth residues, respectively, from the COOH terminus of E2 (Fig. 4). In addition, the kinetics of alanine release by carboxypeptidase Y from radiolabeled E2 and RE2 suggest that they have the same COOH terminus (T. Crowley, personal communication). Taken together, these results establish the COOH terminus of E2 as -Ala-Asn-Ala-COOH, and suggest that RE2 extends to the end of E2.

Properties of RE1 and RE2. The solubility of the glycoprotein roots was examined using two different organic solvent systems (Table 2). Initially we used the two-phase method of Folch et al. (1957) for the extraction of lipids from an aqueous sample using a mixture of chloroform and methanol. RE1 was 4-5 fold more soluble in the organic layer than was RE2. The remainder of both roots was recovered as a precipitate at the aqueous-organic interface. This technique was laborious, and although care was taken to minimize losses, overall recovery was only about 85%. Since both RE1 and RE2 were insoluble in aqueous solutions without detergents, their solubility in a single-phase mixture of acetone, ethanol, and water (5:5:1) was examined. This method was rapid, and recoveries exceeded 95%. RE1 was 3-fold more soluble in this mixture than was RE2. Under the same conditions, the structural proteins were completely insoluble, and the majority of the ^{3}H-palmitate labeled lipids were solubilized from virions (approximately 4% of the palmitate label is covalently associated with E1 and E2, see Fig. 6; Schmidt et al., 1979). This differential solubility of RE1 and RE2 in organic solvents, in addition to being potentially useful as a separation method, suggested that RE1 was more hydrophobic in nature than was RE2. The amino acid compositions of the roots are in agreement with this finding (see Table 1).

Two other potentially useful methods for separating these hydrophobic molecules

Table 2

Organic Solubility of RE1 and RE2[a]

	% soluble in organic phase	
	System A	System B
RE1	77	70
RE2	16	23
Capsid protein	-	<0.5
^{3}H-palmitate Sindbis	-	94

[a]Solvent system A was that of Folch et al. (1957). Solvent system B was acetone:ethanol:water (5:5:1). Extractions were done as described in the Materials and Methods.

are shown in Fig. 5. RE1 and RE2 could be separated using SDS hydroxylapatite chromatography (Moss and Rosenblum, 1972) by elution with a shallow linear gradient from 0.15 M to 0.25 M $NaPO_4$, pH 6.4 (Fig. 5A). RE2 eluted first at 0.207 M $NaPO_4$, followed by RE1 at 0.234 M $NaPO_4$ (these values correspond to the concentration of $NaPO_4$ in the peak fractions). The more hydrophobic root (RE1) although smaller than RE2 binds more tightly to hydroxylapatite. Recovery was nearly quantitative. The two roots could also be separated on the basis of size using gel filtration on Ultrogel AcA 44 in the presence of SDS (Fig. 5B). RE2 eluted first as a sharp peak (recovery 82%), whereas RE1 (recovery 67%) eluted as a broad peak trailing into the included column volume. The broad peak and low recovery of RE1 may have resulted from reversible interactions with the column resin.

Association of palmitate with RE1 and RE2. It has been shown recently that Sindbis glycoproteins have covalently attached fatty acid (Schmidt et al., 1979). E2 appears to have 5-6 molecules, and E1 only about 1-2 molecules. We have investigated the possible association of ^{3}H-palmitate with the glycoprotein roots. Figure 6A compares intact virions or spikeless particles labeled with either ^{35}S-methionine or ^{3}H-palmitate. It is apparent that RE2 contains associated palmitate. RE1 also appears to contain palmitate label although it is poorly resolved from the ^{3}H-palmitate labeled phospholipids migrating at the front. In order to quantitate the fraction of palmitate label found in the roots, preparations of virus labeled with methionine and palmitate were mixed, digested with α-chymotrypsin, and the components of the purified spikeless particles were separated on cylindrical gels (Fig. 6B). The same fraction of the palmitate label was found in RE2 and E2. RE1 also contained palmitate label, and although quantitation was difficult due to the large phospholipid peak at the front, E1 and RE1 had approximately the same proportion of the palmitate label. This demonstrates that the covalently attached lipid is localized in the protease-resistant, membrane-associated region of each glycoprotein.

Fig. 5. Separation of RE1 and RE2 by hydroxylapatite chromatography and gel filtration. Samples of gel purified RE1 and RE2 were adjusted to 1% SDS, 0.1% β-mercaptoethanol and heated to 56°C for 10 minutes. For hydroxylapatite chromatography (panel A), the solution was adjusted to 0.1% SDS and 10 mM sodium phosphate ($NaPO_4$), pH 6.4, and applied to a 0.8 x 0.5 cm column of hydroxylapatite (Moss and Rosenblum, 1972) equilibrated in 10 mM $NaPO_4$, pH 6.4, 0.1% SDS, and 2 mM DTT at 37°C. The column was washed with 5 ml of this buffer, followed by 10 ml of 0.15 M $NaPO_4$, pH 6.4, 0.1% SDS, and 2 mM DTT. The proteins were eluted with a linear gradient from 0.15 M $NaPO_4$ to 0.25 M $NaPO_4$, pH 6.4, in 0.1% SDS, and 2 mM DTT. For separation by gel filtration (panel B), chromatography was performed on a 0.5 x 100 cm column of Ultrogel AcA44 (LKB) equilibrated in 0.1% SDS, 2 mM DTT, and 2 mM Tris-Cl, pH 7.4. The excluded and included column volumes (measured by including blue dextran and phenol red in the sample) are indicated by V_E and V_I, respectively.

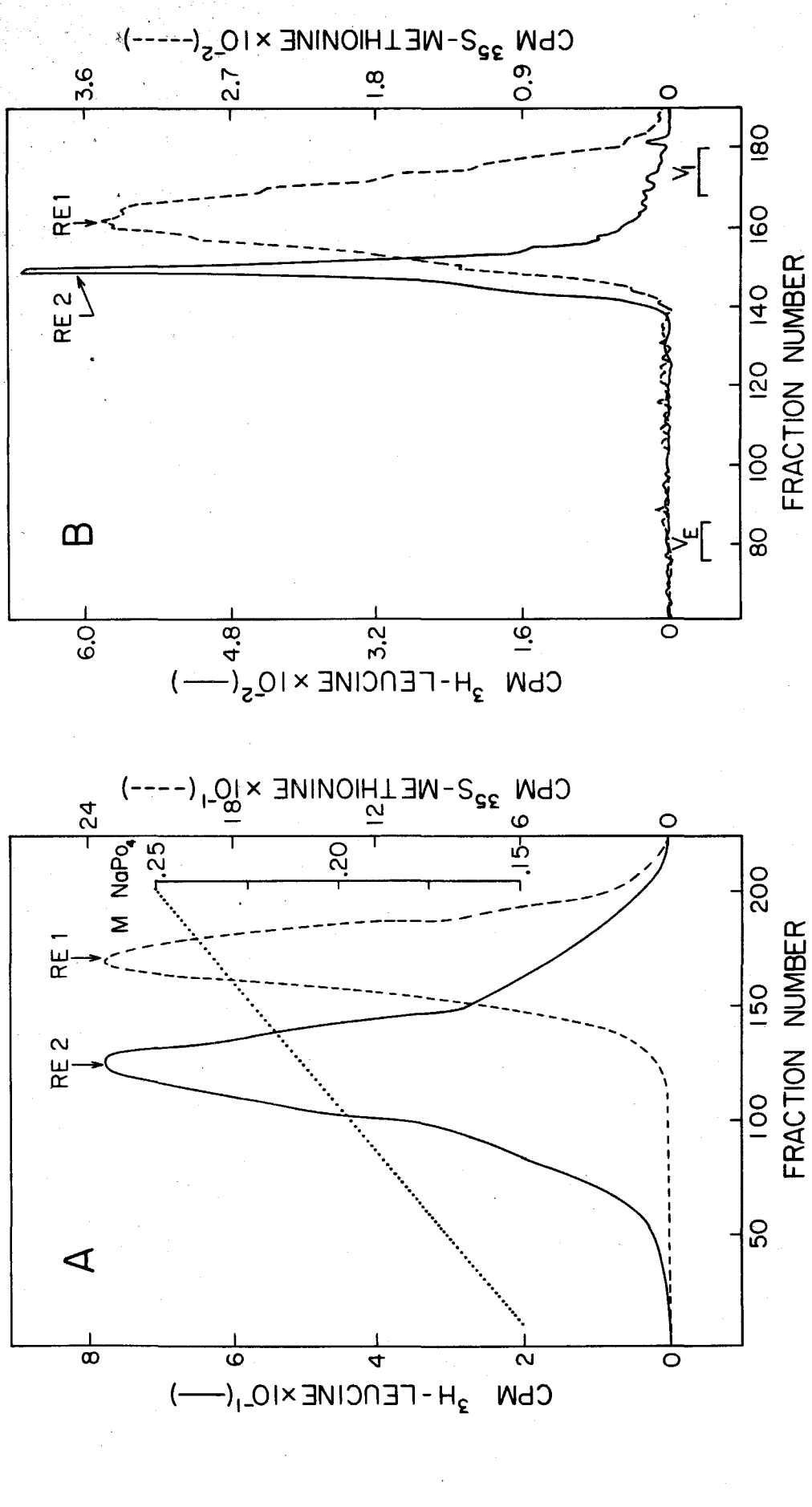
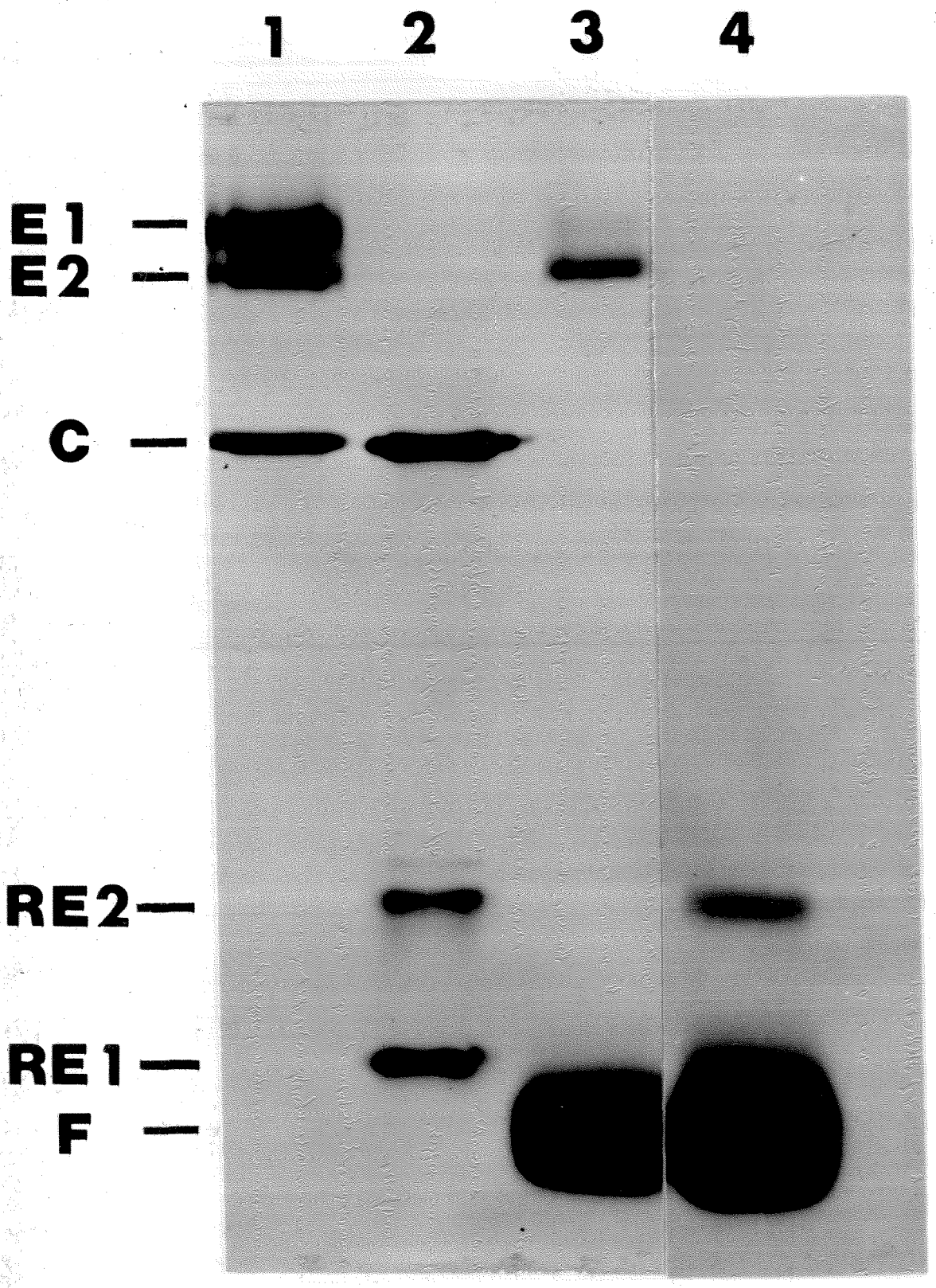
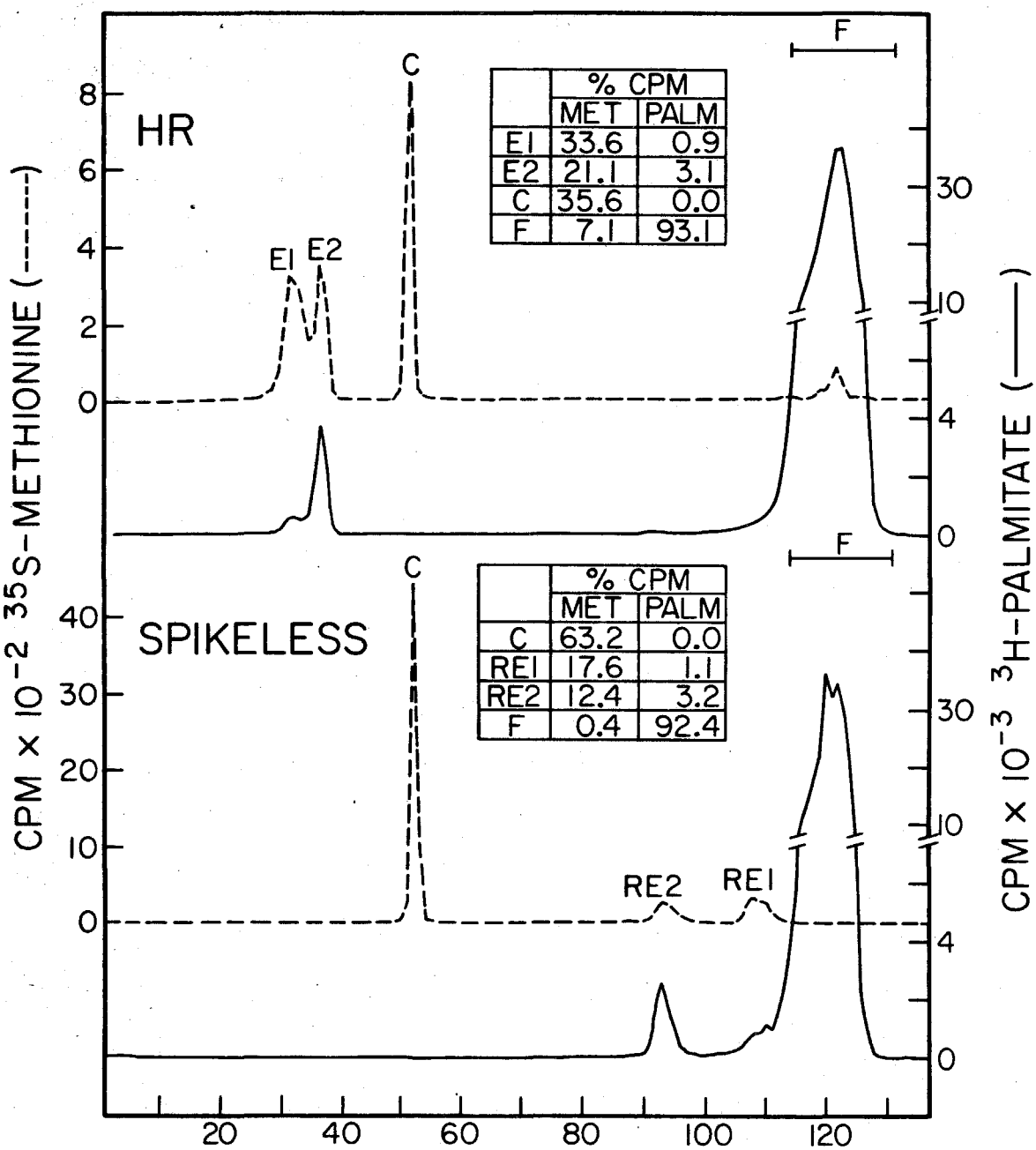

A
RE 1
RE 2
M NaPo4
.25
.20
.15
24
18
12
6
0
8
6
4
2
0
50
100
150
200
FRACTION NUMBER
CPM 3H-LEUCINE ×10-1 (——)
CPM 35S-METHIONINE ×10-1 (----)
B
RE1
RE 2
VI
VE
3.6
2.7
1.8
0.9
0
6.0
4.8
3.2
1.6
0
80
100
120
140
160
180
FRACTION NUMBER
CPM 3H-LEUCINE ×10-2 (——)
CPM 35S-METHIONINE ×10-2 (----)

Fig. 6. Association of palmitate with RE1 and RE2. Purified samples of ^{35}S-methionine or ^{3}H-palmitate labeled Sindbis virus and spikeless particles were prepared as described in the Materials and Methods. In panel A, portions were run on a 12-20% acrylamide exponential gradient gel: lane 1, ^{35}S-methionine labeled Sindbis virus; lane 2, ^{35}S-methionine labeled spikeless particles; lane 3, ^{3}H-palmitate labeled Sindbis virus; lane 4, ^{3}H-palmitate labeled spikeless particles. In panel B, portions of the ^{35}S-methionine and ^{3}H-palmitate labeled samples were mixed, run on 20% acrylamide cylindrical gels, and the peaks of radioactivity quantitated (given at the percent of the total cpm recovered) as described in the Materials and Methods. Electrophoresis was from left to right.


1
2
3
4
E1
E2
C
RE2
RE1
F


HR
E1
E2
C
F
% CPM
MET PALM
E1 33.6 0.9
E2 21.1 3.1
C 35.6 0.0
F 7.1 93.1
SPIKELESS
C
RE2
RE1
F
% CPM
MET PALM
C 63.2 0.0
RE1 17.6 1.1
RE2 12.4 3.2
F 0.4 92.4
CPM x 10-2 35S-METHIONINE (------)
CPM x 10-3 3H-PALMITATE (——)
0
2
4
6
8
10
20
30
40
60
80
100
120

DISCUSSION

Several lines of evidence indicate that the E1 and E2 glycoproteins of Sindbis virus are transmembrane proteins. α-Chymotrypsin digestion of intact virions yields a single membrane-protected peptide derived from each glycoprotein. Both of these roots contain long stretches of hydrophobic amino acids near their NH_2 termini of sufficient length to span the bilayer in an α-helix. E2 and its precursor PE2 have been shown directly to span the membrane by protease treatment of microsomes from both infected cell lysates (Wirth et al., 1977; Ziemiecki et al., 1980) and heterologous cell free systems (Garoff et al., 1978; Bonatti et al., 1979). Assuming that E2 in mature virions has a similar conformation with respect to the lipid bilayer, then the observation that protease treatment of E2 decreases its size by about 30 amino acids implies that at least one protease sensitive site exists shortly after the first hydrophobic domain found in the RE2 sequence (see Fig. 4). Whether the entire 33 amino acid sequence COOH-terminal to this transmembrane region is accessible to protease or is partially protected by the lipid bilayer is currently under investigation.

In contrast to E2, the apparent molecular weight of E1 is not changed by protease treatment of infected cell microsomes. This result is expected since the presumptive cytoplasmic COOH terminus of E1 would contain only a few resides, none of which are good candidates for α-chymotrypsin digestion. Although no direct evidence for the transmembrane conformation of E1 exists, the similarity of its structure and mode of synthesis to other simple transmembrane proteins suggest that it probably spans the bilayer. The glycoprotein precursors, PE2 and E1, are cotranslationally inserted into the lumen of the rough endoplasmic reticulum (Garoff et al., 1978; Bonatti et al., 1979). This process is initiated by a hydrophobic amino-terminal extension of the nascent polypeptide, or signal sequence (probably the 6K protein in the case of E1) (for reviews, see Blobel et al., 1979; Davis and Tai, 1980; Wickner,

1980), leading to the vectorial transport of polypeptide across the membrane. For the limited number of proteins where data exists, integral membrane proteins synthesized in this fashion are usually anchored in the bilayer by means of a short COOH-terminal hydrophobic peptide (see below, and Table 3). In such cases, the majority of the polypeptide mass, including the NH_2 terminus, can be removed by protease treatment from the extracytoplasmic side. The COOH-terminal anchors of these proteins typically have a cluster of basic residues immediately following the transmembrane segment on the cytoplasmic side of the bilayer. The E1 glycoprotein of Sindbis virus contains a membrane-protected COOH-terminal sequence of 33 uncharged residues followed by two arginines, and by analogy to other well characterized transmembrane proteins, probably spans the bilayer (see below and Table 3). Since many membrane proteins may have only short cytoplasmic extensions, the use of lipid insoluble protein labeling reagents on isolated microsomes may prove more universally applicable than protease treatment for demonstrating their transmembrane character.

As mentioned above, several other transmembrane glycoproteins have COOH-terminal hydrophobic segments and are summarized in Table 3. The hydrophobic tails serve to anchor such glycoproteins in the bilayer and limited proteolysis often solubilizes a large and sometimes functional portion of the protein without the use of detergents (Skehel and Waterfield, 1975; Melcher et al., 1975; Ozols et al., 1976). The importance of this hydrophobic segment is illustrated by the heavy chain of IgM which occurs in two forms; one of which is membrane bound and another which is secreted (Rogers et al., 1980; Kehry et al., 1980). The major structural difference between these two polypeptides is the presence of a COOH-terminal hydrophobic segment in the membrane form which is absent in the secreted form. In general, the transmembrane domains of these proteins are short (from 20-33 amino acids), contain predominantly hydrophobic amino acid residues (including Ser and Thr), and can be roughly predicted based on the exclusion of Asp, Glu, Asn, Gln, His, Lys,

Table 3

Structure of Transmembrane Sequences

Protein[a]	Number of residues in[b] hydrophobic sequence	Number of residues COOH-terminal to hydrophobic sequence	Nearby charged residues[c]		Reference[d]
			NH_2-terminal	COOH-terminal	
Sindbis E1	33	2	K(1)	R(1-2)	1
Sindbis E2	26	33	H(1),R(2),H(3)	K(1),R(3-4),E(5)	1
SFV E1	24	2	K(1)	R(1-2)	2
SFV E2	28	31	None	R(1-2),K(3)	2
WEE E1	?	2	?	R(1-2)	3
VSV G	20	29	K(1)	R(1),H(5)	4
Influenza HA	26	11	D(1),K(2)	K(1 or 2)[e]	5,6
Adenovirus E3/16	23	15	E(2),K(5)	K(1),K(3),R(5)	7
Glycophorin A	23	36	E(1),E(3)	R(1-2),K(5)	8
HLA-A, HLA-B	?	30	?	R(102),K(3)	9
H-2K[b]	24	39	None	K(1),R(3-5)	10
IgM, μM	25	3	E(2),E(5)	K(1),K(3)	11

Table 3 (continued)

[a]Abbreviations: WEE = Western equine encephalitis virus, VSV = vesicular stomatitis virus.

[b]Number of consecutive, unchanged amino acids excluding Arg, Asn, Asp, Gln, Glu, His, Lys, and Pro.

[c]Charged residues (or His) within five residues of the hydrophobic sequence. The numbers are the positions from the end of the uncharged sequence. D = Asp, E= Glu, H= His, K = Lys, R = Arg.

[d]References: (1) Rice and Strauss (1981); (2) Garoff et al. (1980a); (3) J. Ou, personal communication; (4) Rose et al. (1980); (5) Jou et al. (1980); (6) Porter et al. (1980); (7) Persson et al. (1980); (8) Tomita et al. (1978); (9) Robb et al. (1978); (10) Uehara et al. (1981); (11) Rogers et al. (1980).

[e]Two strains.

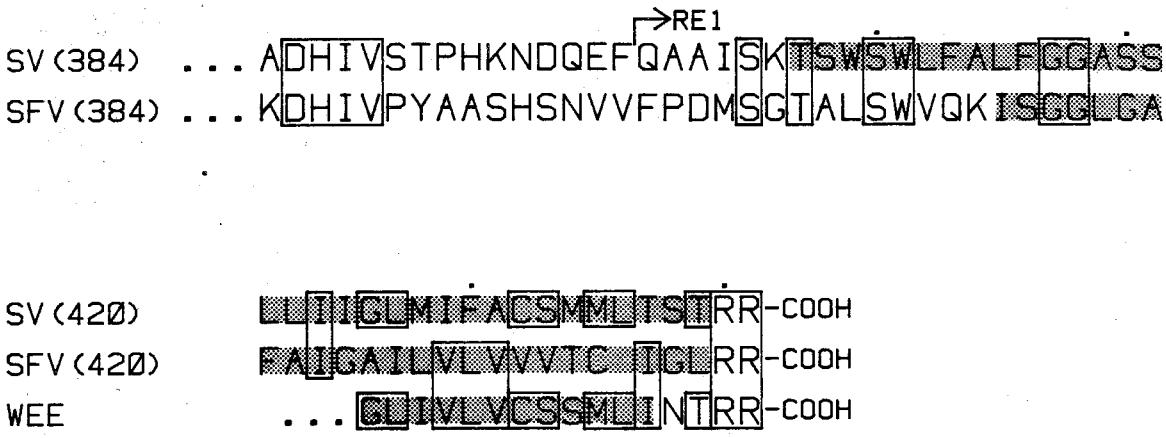
Arg, and Pro. The precise definition of the residues which are actually in contact with hydrophobic interior of the lipid bilayer cannot be deduced from the sequences of these protease-resistant fragments and awaits further investigation perhaps using lipid soluble probes (Goldman et. al., 1979). This difficulty is illustrated by comparison of the Sindbis and SFV E1 glycoproteins. The structural polypeptide sequences of these two closely related alphaviruses can be easily aligned and are 47% homologous with another 12% conservative substitutions (Rice and Strauss, 1981). However, by the above criteria, the putative transmembrane segment in Sindbis E1 is 8 residues longer than that of SFV (Fig. 7). While this additional length may reflect an actual structural difference between these two viruses in the transmembrane region, it could also be due to the inaccuracy of this method for predicting the exact location of transmembrane segments.

Another interesting feature of these transmembrane segments is the presence of clusters of charged residues at one or both ends of the hydrophobic sequence. Basic residues are always found on the cytoplasmic side whereas either acidic or basic residues are usually found near the extracytoplasmic, NH_2-terminal boundary. It has been proposed that the cluster of basic amino acids COOH-terminal to the hydrophobic segment functions as a "stop transfer" signal during the cotranslational insertion and translocation of this class of membrane proteins across the membrane of the rough endoplasmic reticulum (Blobel et al., 1979). The requirement for charged amino acids on the extracytoplasmic side appears to be less stringent than on the cytoplasmic side (see Table 3). It is also of interest to note the COOH-terminal cytoplasmic segments of these proteins are short, containing less than 30-40 amino acids (Table 3).

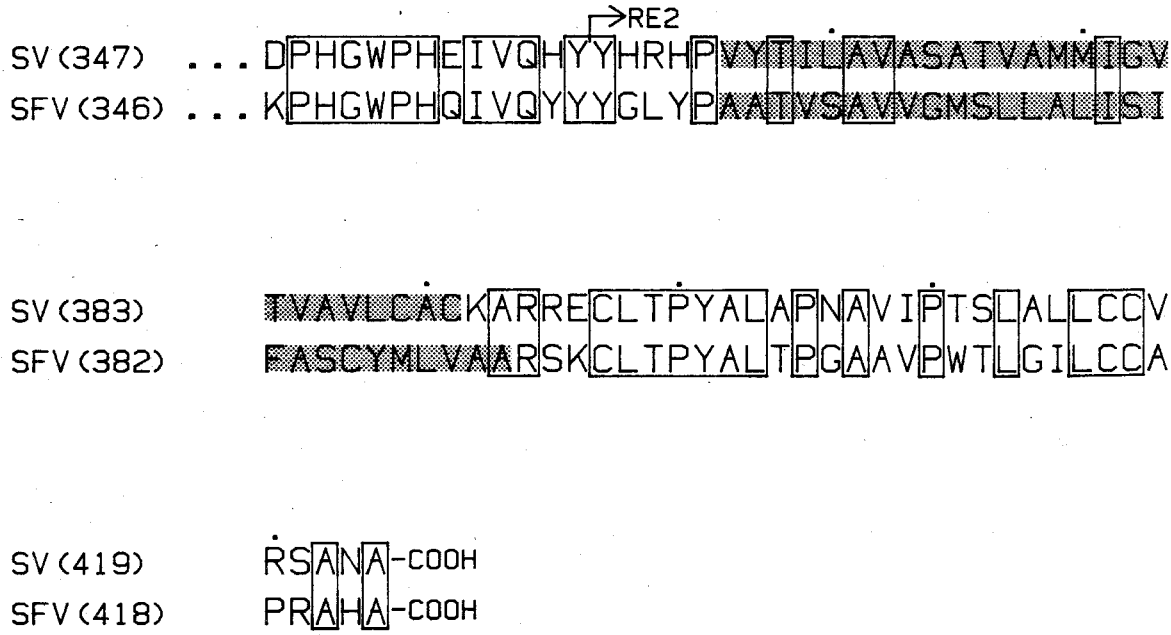
Other membrane proteins such as bacteriorhodopsin (Ovchinnikov et al., 1979; Engelman et al., 1980) and the Band 3 polypeptide in the human erythrocyte membrane (Rao, 1979) span the lipid bilayer several times. Bacteriorhodopsin appears to contain

Fig. 7. Comparison of the COOH termini of alphavirus glycoproteins. The COOH-terminal glycoprotein sequences of Sindbis virus (SV) (Rice and Strauss, 1981; see also Fig. 4), SFV (Garoff et al., 1980a), and Western equine encephalitis virus (WEE) (J. Ou, personal communication) are aligned to maximize homology using protein and nucleic acid sequence data. Homologous residues are enclosed in boxes. Inclusion of the data for WEE suggests a different alignment of the Sindbis virus and SFV sequences in the region of E1 shown than previously presented (Rice and Strauss, 1981). The single letter amino acid code is used: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. The numbers indicate the position from the amino terminus of the glycoprotein. The amino termini of the roots of Sindbis virus E1 and E2 (see Fig. 4) are marked by arrows. Putative transmembrane segments as defined in the text (see Discussion) and Table 3 are shaded.

COOH TERMINUS OF E1



COOH TERMINUS OF E2



several charged residues in the lipid bilayer, and a stable intramembrane structure can be constructed in which seven transmembrane helices are arranged such that these charged residues can form ion pairs shielded from the lipid environment (Engleman et al., 1980). Similarly, the major outer membrane protein I of Escherichia coli B/r, which serves as a transmembrane pore (probably consisting of three polypeptides) allowing the passage of charged solutes through the outer bacterial membrane, does not contain any sequences of nonpolar amino acids longer than 11 residues (Chen et al., 1979). In contrast to the transmembrane segments described in Table 3 which are simple, uninterrupted sequences of nonpolar amino acids serving as lipophilic membrane anchors, these more complicated transmembrane structures are undoubtedly related to their additional function in the passage of charged solutes across membranes. It would be expected that such specialized transmembrane segments would be highly conserved between functionally similar molecules derived from a common evolutionary origin. On the other hand, transmembrane segments anchoring proteins in the lipid bilayer would only be constrained to retain their hydrophobic character and structural properties (such as the formation of an α-helix). This is illustrated by a comparison of the presumptive transmembrane segments of Sindbis virus and SFV E1 and E2 (Fig. 7). The putative transmembrane domains are only 15-16% homologous in contrast to the overall homology of 47% between the structural proteins. All of the substitutions involve nonpolar amino acids, and cysteine residues which are highly conserved in other regions of E1 and E2 (note the extreme COOH-terminal region of E2 shown in Fig. 7) are not conserved in their transmembrane segments. Although only limited data are currently available for Western equine encephalitis virus, it is of interest to note that at the COOH-terminal end of E1 only the two arginines are common to all three of these closely related alphaviruses (Fig. 7).

In addition to stretches of COOH-terminal hydrophobic amino acids, Sindbis virus E1 and E2 (Schmidt et al., 1979) and vesicular stomatitis virus glycoprotein G

(Schmidt and Schlesinger, 1979) contain covalently attached fatty acids. In the case of Sindbis virus, we have shown that the palmitate attached to the glycoproteins is localized in their membrane-associated hydrophobic roots. Although the function of these fatty acids remains unknown, the attachment of such hydrophobic groups to transmembrane polypeptide segments might enhance their solubility in the lipophilic environment of the bilayer. However, it should be mentioned that the covalent attachment of palmitate is a post-translational event probably occuring in the Golgi 10-20 minutes after the polypeptides have been asymmetrically inserted into the membrane of the rough endoplasmic reticulum (Schmidt and Schlesinger, 1980) and may be important in other aspects of virus structure and maturation.

The cytoplasmic portions of the Sindbis glycoproteins are of great interest since they interact specifically with the nucleocapsid localized in the cytoplasm leading to the budding of mature virions containing exclusively viral glycoproteins (Strauss, 1978). As mentioned above, the transmembrane segments of Sindbis virus and SFV E1 and E2 are conserved only with respect to their hydrophobic character, and the cytoplasmic COOH-terminal end of E1 (assuming that it completely spans the bilayer) would consist of only a few residues. In contrast to its COOH-terminal transmembrane segment, E2 contains a second stretch of 23 nonpolar amino acids which are 61% homologous between Sindbis virus and SFV, with another 26% of the residues being conservative substitutions (Fig. 7). This region contains three conserved prolines as well as three conserved cysteines which are not typical of the simple transmembrane segments shown in Table 3, and suggest that this segment does not function as a simple lipophilic membrane anchor. Whether this region of E2 loops back into the bilayer or is exposed on the cytoplasmic side of the plasma membrane is at present unknown. However, its highly conserved nature and COOH-terminal location in E2 make it a likely candidate for specific interaction with the nucleocapsid, and we are presently investigating this possibility using Sindbis ts mutants (Strauss

and Strauss, 1980). If this premise is true, it implies that this interaction may be at least partially hydrophobic in nature. This is supported by the observation that treatment of virions with non-ionic detergents such as Triton X-100 disrupts this interaction (Helenius and Söderlund, 1973).

The amino acid compositions and molecular weights for the Sindbis virus structural proteins are now accurately known (Bell et al., 1979; Rice and Strauss, 1981), and it has been possible to directly determine the stoichiometry of E1, E2, and C using the incorporation of different radiolabeled amino acids into virions. The data presented here clearly show that RE1, RE2, and the capsid protein (and therefore E1, E2, and C) are present in roughly equimolar amounts. This finding has been previously reported for Sindbis virus and SFV (Schlesinger et al., 1972; Garoff et al., 1974). In addition, cross-linking experiments have shown that spike glycoproteins E1 and E2 are associated as heterodimers in the viral membrane and after disruption with Triton X-100 (Ziemiecki and Garoff, 1978; Rice and Strauss, in preparation). This assures that the interaction of either glycoprotein (presumably E2) in a spike with each capsid protein subunit of the nucleocapsid will produce budded virions containing equimolar amounts of the structural proteins.

ACKNOWLEDGEMENTS

We would like to thank E. Lenches for preparing monolayers of chicken embryo-fibroblasts, L. Hood for his generosity in lending us equipment for amino acid analysis and protein sequence determination, and W. T. Welch and B. M. Sefton for allowing us to quote their unpublished results. This work was supported by grants GM06965 and AI10793 from NIH, by grant PCM 80-22830 from the National Science Foundation, and by Biomedical Research Support grant 507RR07003 from NIH. C.M.R. was supported by Training Grant GM00086 from NIH, and by a fellowship from the California Foundation for Biochemical Research.

REFERENCES

Bell, J. R., Hunkapiller, M. W., Hood, L. E., and Strauss, J. H. (1978) Amino-terminal sequence analysis of the structural proteins of Sindbis virus. Proc. Natl. Acad. Sci. USA **75**, 2722-2726.

Bell, J. R., Strauss, E. G., and Strauss, J. H. (1979) Purification and amino acid compositions of the structural proteins of Sindbis virus. Virology **97**, 287-294.

Blobel, G., Walter, P., Chang, C., Goldman, B. M., Erickson, A. H., and Lingappa, V. R. (1979). Translocation of proteins across membranes: the signal hypothesis and beyond. In "Symposia of the Society for Experimental Biology," No. 33, pp. 9-36.

Bonatti, S., Cancedda, R., and Blobel, G. (1979) Membrane biogenesis: in vitro cleavage, core glycosylation, and integration into microsomal membranes of Sindbis virus glycoproteins. J. Cell Biol. **80**, 219-224.

Bonner, W. M., and Laskey, R. A. (1974) A film detection method for tritium-labeled proteins and nucleic acids in polyacrylamide gels. Eur. J. Biochem. **46**, 83-88.

Cancedda, R., Villa-Komaroff, L., Lodish, H. F., and Schlesinger, M. (1975) Initiation sites for translation of Sindbis virus 42S and 26S messenger RNAs. Cell **6**, 215-222.

Chen, R., Kramer, L., Schmidmayr, W., and Henning, U. (1979) Primary structure of the major outer membrane protein I of Escherichia coli B/r. Proc. Natl. Acad. Sci. USA **76**, 5014-5017.

Compans, R. W. (1971) Location of the glycoproteins in the membrane of Sindbis virus. Nature New Biol. **229**, 114-116.

Davis, B. D., and Tai, P.-C. (1980). The mechanism of protein secretion across membranes. Nature (London) **283**, 433-438.

Engelman, D. M., Henderson, R., McLachlan, A. D., and Wallace, B. A. (1980) Path of the polypeptide of bacteriorhodopsin. Proc. Natl. Acad. Sci. USA **77**, 2023-2027.

Erwin, C., and Brown, D. T. (1980). Intracellular distribution of Sindbis virus membrane proteins in BHK-21 cells infected with wild-type virus and maturation-defective mutants. J. Virol. **36**, 775-786.

Folch, J., Lees, M., and Sloane-Stanley, G. H. (1957) A simple method for the isolation and purification of total lipids from animal tissues. J. Biol. Chem. **226**, 497-509.

Garoff, H., and Simons, K. (1974) Location of the spike glycoproteins in the Semliki Forest virus membrane. Proc. Nat. Acad. Sci. USA **71**, 3988-3992.

Garoff, H., Simons, K., and Renkonen, O. (1974) Isolation and characterization of the membrane proteins of Semliki Forest virus. Virology **61**, 493-504.

Garoff, H., and Söderlund, H. (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.

Garoff, H., Simons, K., and Dobberstein, B. (1978) Assembly of the Semliki Forest virus membrane glycoproteins in the membrane of the endoplasmic reticulum in vitro. J. Mol. Biol. **124**, 587-600.

Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H., and Delius, H. (1980a) Nucleotide sequence of the cDNA coding for the Semliki Forest virus membrane glycoproteins. Nature (London) **288**, 236-241.

Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H., and Delius, H. (1980b) The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino terminal region. Proc. Natl. Acad. Sci. USA **77**, 6376-6380.

Goldman, D. W., Pober, J. S., White, J., and Bayley, H. (1979). Selective labeling of the hydrophobic segments of intrinsic membrane proteins with a lipophilic photogenerated carbene. Nature (London) **280**, 841-843.

Helenius, A., and Söderlund, H. (1973) Stepwise dissociation of the Semliki Forest virus membrane with Triton X-100. Biochem. Biophys. Acta **307**, 287-300.

Hirs, C. H. W. (1967) Performic acid oxidation. in Methods of Enzymology. ed. Hirs, C. H. W. (Academic Press, New York) Vol. 11, pp. 197-199.

Hunkapiller, M. W., and Hood, L. E. (1978) Direct microsequence analysis of polypeptides using an improved sequenator, a nonprotein carrier (Polybrene), and high pressure liquid chromatography. Biochemistry **17**, 2124-2133.

Hunkapiller, M. W., and Hood, L. E. (1980) A new protein sequenator with improved sensitivity. Science **207**, 523-525.

Johnson, N., Hunkapiller, M., and Hood, L. (1979). Analysis of phenylthiohydantoin amino acids by high-performance liquid chromatography on DuPont Zorbax cyanopropylsilane columns. Anal. Biochem. **100**, 335-338.

Jou, W. M., Verhoeyen, M., Devon, R., Saman, E., Fang, R., Huylebroeck, D., Fiers, W., Threlfall, G., Barber, C., Carey, N., and Emtage, S. (1980). Complete structure of the hemagglutinin gene from the human influenza A/Victoria/3/75 (H3N2) strain as determined from cloned cDNA. Cell **19**, 683-696.

Katz, F. N., Rothman, J. E., Lingappa, V. R., Blobel, G., and Lodish, H. F. (1977) Membrane assembly in vitro: synthesis, glycosylation and asymmetric insertion of a transmembrane protein. Proc. Natl. Acad. Sci. USA **74**, 3278-3282.

Kehry, M., Ewald, S., Douglas, R., Sibley, C., Raschke, W., Fambrough, D., and Hood, L. (1980) The immunoglobulin μ chain of membrane-bound and secreted IgM molecules differ in their C-terminal segments. Cell **21**, 393-406.

Laemmli, U. K. (1970) Cleavage of the structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**, 680-685.

Laskey, R. A., Mills, A. D. (1975) Quantitative film detection of ^{3}H and ^{14}C in polyacrylamide gels by fluorography. Eur. J. Biochem. **56**, 335-341.

McMillan, M., Cecka, J. M., Hood, L., Murphy, D. B., McDevitt, H. O. (1979) Peptide map analysis of murine Ia antigens of the I-E subregion using HPLC. Nature (London) **277**, 663-664.

Melcher, U., Eidels, L., and Uhr, J. W. (1975) Are immunoglobulins integral membrane proteins? Nature (London) **258**, 434-435.

Moss, B., and Rosenblum, E. N. (1972) Hydroxylapatite chromatography of protein-sodium dodecyl sulfate complexes. J. Biol. Chem. **247**, 5194-5198.

Neal, M. W., and Florini, J. R. (1973) A rapid method for desalting small volumes of solution. Analyt. Biochem. **55**, 328-330.

Ovchinnikov, Yu. a., Abdulaev, N. G., Feigina, M. Yu., Kiselev, A. V., and Lobanov, N. A. (1979) The structural basis of the functioning of bacteriorhodopsin: an overview. FEBS Lett. **100**, 219-224.

Ozols, J., Gerard, C., and Nobrega, F. G. (1976) Proteolytic cleavage of horse liver cytochrome b5: primary structure of the heme-containing moiety. J. Biol. Chem. **251**, 6767-6774.

Palade, G. (1975) Intracellular aspects of the process of protein synthesis. Science **189**, 347-358.

Persson, H., Jörnvall, H., and Zabielski, J. (1980). Multiple mRNA species for the precursor to an adenovirus-encoded glycoprotein: identification and structure of the signal sequence. Proc. Natl. Acad. Sci. USA **77**, 6349-6353.

Pierce, J. S., Strauss, E. G., and Strauss, J. H. (1974) Effect of ionic strength on the binding of Sindbis virus to chick cells. J. Virol. **13**, 1030-1036.

Piszkiewicz, D., Landon, M., and Smith, E. L. (1970). Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations. Biochem. Biophys. Res. Comm. **40**, 1173-1178.

Porter, A. G., Barber, C., Carey, N. H., Hallewell, R. A., Threlfall, G., and Emtage, J. S. (1979). Complete nucleotide sequence of an influenza virus hemagglutinin gene from cloned DNA. Nature (London) **282**, 471-477.

Rao, A. (1979) Disposition of the Band 3 polypeptide in the human erythrocyte membrane. J. Biol. Chem. **254**, 3503-3511.

Rice, C. M., and Strauss, J. H. (1981). Nucleotide sequence of the 26S mRNA of Sindbis virus and deduced sequence of the encoded virus structural proteins. Proc. Natl. Acad. Sci. USA, in press.

Robb, R. J., Terhorst, C., and Strominger, J. L. (1978). Sequence of the COOH-terminal hydrophobic region of histocompatibility antigens HLA-A2 and HLA-B7. J. Biol. Chem. **253**, 5319-5324.

Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., and Wall, R. (1980) Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin μ chains. Cell **20**, 303-312.

Rose, J. K., Welch, W. J., Sefton, B. M., Esch, F. S., and Ling, N. C. (1980) Vesicular stomatitis virus glycoprotein is anchored in the viral membrane by a hydrophobic domain near the COOH-terminus. Proc. Natl. Acad. Sci. USA **77**, 3884-3888.

Schlesinger, M. J., Schlesinger, S., and Burge, B. W. (1972). Identification of a second glycoprotein in Sindbis virus. Virology **47**, 539-541.

Schlesinger, M. J., and Kääriäinen, L. (1980) Translation and processing of alphavirus protein. In The Togaviruses, Schlesinger, R. W., ed. (Academic Press, New York) pp. 371-392.

Schmidt, M. F. G., Bracha, M., and Schlesinger, M. J. (1979) Evidence for the covalent attachment of fatty acids to Sindbis virus glycoproteins. Proc. Natl. Acad. Sci. USA **76**, 1687-1691.

Schmidt, M. F. G., and Schlesinger, M. J. (1980) Relation of fatty acid attachment to the translation and maturation of vesicular stomatitus virus and Sindbis virus membrane glycoproteins. J. Biol. Chem. **255**, 3334-3339.

Sefton, B. M. (1977) Immediate glycosylation of Sindbis virus membrane proteins. Cell **10**, 659-668.

Skehel, J. J., and Waterfield, M. D. (1975) Studies on the primary structure of the influenza virus hemagglutinin. Proc. Natl. Acad. Sci. USA **72**, 93-97.

Strauss, E. G. (1978) Mutants of Sindbis virus. III. Host polypeptides present in purified HR and ts103 virus particles. J. Virol. **28**, 266-274.

Strauss, E. G., and Strauss, J. H. (1980). "Mutants of alphaviruses: genetics and physiology" in The Togaviruses, Schlesinger, R. W., ed. (Academic Press, New York) pp. 393-426.

Strauss, J. H., and Strauss, E. G. (1977) Togaviruses In The Molecular Biology of Animal Viruses, Nayak, D., ed. Vol. 1, pp. 111-166.

Tomita, M., Furthmayr, H., and Marchesi, V. T. (1978). Primary structure of human erythrocyte glycophorin A. Isolation and characterization of peptides and complete amino acid sequence. Biochemistry **17**, 4756-4770.

Uehara, H., Coligan, J. E., and Nathenson, S. G. (1981). Isolation and sequence analysis of membrane and hydrophilic segments of the H-2K^{b} alloantigen. Biochemistry, in press.

Uterman, G., and Simons, K. (1974) Studies on the amphipathic nature of the membrane proteins of Semliki Forest virus. J. Mol. Biol. **85**, 569-587.

Welch, W. J., and Sefton, B. M. (1979) Two small virus-specific polypeptides are produced during infection with Sindbis virus. J. Virol. **29**, 1186-1195.

Wickner, W. (1980). Assembly of proteins into membranes. Science **210**, 861-868.

Wirth, D. F., Katz, F., Small, B., and Lodish, H. F. (1977) How a single Sindbis virus mRNA directs the synthesis of one soluble and two integral membrane glycoproteins. Cell **10**, 253-263.

Ziemiecki, A., and Garoff, H. (1978) Subunit composition of the membrane glycoprotein complex of Semliki Forest virus. J. Mol. Biol. **122**, 259-269.

Ziemiecki, A., Garoff, H., and Simons, K. (1980). Formation of the Semliki Forest virus membrane glycoprotein complexes in the infected cell. J. gen. Virol. **50**, 111-123.

CHAPTER 5

Association of Sindbis virion glycoproteins and their precursors

CHARLES M. RICE and JAMES H. STRAUSS

ABSTRACT We have studied the association of the Sindbis virus glycoproteins in mature virions and infected cells. The glycoproteins were cross-linked with bifunctional amino-reactive imidates (11 Å cross-linking distance) which could be subsequently cleaved by reduction. Using monospecific rabbit antisera against each viral envelope glycoprotein it was found that >90% of the cross-linked glycoprotein dimers from intact virions or virions solubilized with Triton X-100 prior to cross-linking were heterodimers of E1 and E2. The pattern of cross-linked glycoproteins from intact virions as well as infected cells suggested that three E1-E2 dimers may be associated to form a hexameric subunit. Cross-linking of pulse-labeled infected monolayers showed that PE2 was cross-linked to E1 less efficiently than was E2, suggesting that if PE2 and E1 are associated as a complex in infected cells then their conformation with respect to reactive amino groups is distinct from that of the mature virion glycoproteins. ts mutants of Sindbis virus in the complementation groups corresponding to E1 and PE2 fail to cleave PE2 at the nonpermissive temperature (40°C). In monolayers infected with these mutants or a heat resistant variant of Sindbis virus, the glycoprotein precursors synthesized at the elevated temperature were readily cross-linked into large aggregates, indicating a temperature-sensitive tendency for the proteins to aggregate.

INTRODUCTION

Alphaviruses are simple enveloped animal viruses useful in the study of glycoprotein biosynthesis. Sindbis virus and the closely related Semliki Forest virus (SFV) contain two or three membrane glycoproteins (E1, E2 and in the case of SFV, E3) anchored in a lipid bilayer of host cell origin which surrounds an iscosahedral nucleocapsid consisting of approximately 240 capsid protein (C) subunits and a genomic 49S RNA (Strauss and Straüss, 1977). The virion structural proteins are encoded by a 3' terminal subgenomic 26S RNA (Simmons and Strauss, 1974; Ou et al., 1981) with a single initiation site for protein synthesis (Cancedda et al., 1975). The genes are translated on membrane-bound polyribosomes (Wirth et al., 1977) in the order 5'-C-E3-E2-6K-E1-3' (Schlesinger and Kääriäinen, 1980; Garoff et al., 1980a, b; Rice and Strauss, 1981), as a continuous polypeptide which is processed by proteolytic cleavage. The capsid protein is cleaved while nascent and associates with the 49S RNA in the cytoplasm to form the nucleocapsid. The glycoproteins are cotranslationally inserted into the lumen of the rough endoplasmic reticulum (Garoff et al., 1978; Bonatti et al., 1979) and core glycosylated (Sefton, 1977). These mannose-rich oligosaccharide units are subsequently modified (presumably in the smooth endoplasmic reticulum) and about 20 min after their synthesis the glycoproteins can be found in the host cell plasma membrane. Both E1 and E2 can be found in mature virions beginning at about 20-30 min following their synthesis, and the cleavage of PE2 (the precursor to E2) which also occurs at this time, is necessary for the final steps in virus maturation (Strauss and Strauss, 1977). About 30 COOH-terminal amino acids can be removed from E2 (Ziemiecki et al., 1980) and PE2 (Wirth et al., 1977; Garoff and Söderlund, 1978; Bonatti et al., 1979) by proteolysis of microsomes from infected cells or heterologous cell free systems. Glycoprotein E1, although unaffected by protease treatment, is probably also a transmembrane protein containing at most a few amino acids exposed on the cytoplasmic side of the bilayer (Rice and Strauss, 1981; Rice et al., in preparation).

The interaction of the nucleocapsid with the cytoplasmic portion of the glycoprotein(s) (probably only PE2 or E2) results in virus budding through the host cell plasma membrane, producing a mature virion containing equimolar amounts of each structural protein (Schlesinger et al., 1972; Garoff et al., 1974; Rice et al., in preparation). This interaction is highly specific and host cell glycoproteins are rigorously excluded from released virus particles (Strauss, 1978).

The equimolar stoichiometry of the structural proteins and coincident appearance of the glycoproteins in mature virions (Schlesinger and Schlesinger, 1972) led to the early suggestion that PE2 and E1 might exist in a stable complex in infected cells. It was later found that ts mutants in complementation groups corresponding to either E1 or PE2 are defective in the PE2 cleavage at the restrictive temperature (Bracha and Schlesinger, 1976; Jones et al., 1977; Smith and Brown, 1977), and that this cleavage was inhibited by antiserum to either E1 or E2 (Bracha and Schlesinger, 1976; Jones et al., 1977). In the case of SFV, it has been shown that E1 and E2 can be chemically cross-linked into a heterodimer both in intact virions and after solubilization with Triton X-100 (Ziemiecki and Garoff, 1978). Using antisera specific for each of the virion structural proteins and bifunctional cross-linking reagents, we have investigated the possiblities that an association exists between the E1 and E2 glycoproteins of mature Sindbis virions, and that E1 and PE2 form a similar complex in infected cells.

MATERIALS AND METHODS

Cells and Virus Strains. Virus was grown in confluent monolayers of either primary chicken embryo fibroblasts (Pierce et al., 1974) or hamster cells (BHK-21). The heat-resistant (large plaque) strain of Sindbis virus (HR) and temperature-sensitive mutants ts20 and ts23 of Burge and Pfefferkorn (1966) were generously furnished by Dr. B.W. Burge. All strains have recently been plaque purified in this laboratory. Semliki

Forest virus (SFV) was obtained from Dr. Judith Levin and passaged in primary chicken embryo fibroblasts.

Virus Growth and Purification. Radiolabeled Sindbis virus (HR strain) grown in roller bottles (750 cm^2) of confluent primary chicken embryo fibroblasts (CEF) was prepared by the salt-reversal method (Pierce et al., 1974). The high-salt released virus was purified by rate zonal contrifugation and isopycnic banding as previously described (Bell et al., 1979) except that bovine serum albumin (BSA) (Sigma, fraction V) at a concentration of 200 μg/ml was included in all gradients, and 0.05 M triethanolamine, pH 8.5, was substituted for 0.05 M Tris-Cl, pH 7.4, during isopycnic banding.

Radiolabeled HR and SFV were also grown in BHK-21 cells in the presence of normal concentrations of NaCl (0.116 M). This virus was purified from the culture medium harvested at 12 hr after infection (37°C) by a single isopycnic centrifugation.

For preliminary studies, milligram quantities of CEF grown HR and SFV were purified after polyethylene glycol precipitation by the method of Bell et al. (1979) or Kääriäinen et al. (1969).

Cross-linking of Isolated Virions. Stock solutions of dithiobis (succinimidyl proprionate) (DTSP) (Pierce Chemicals) dissolved in 100% dimethyl sulfoxide (DMSO) or dimethyl suberimidate (DMS) (Pierce Chemicals) dissolved in 0.2 M triethanolamine, pH 8.5, were made immediately before use. For cross-linking with DTSP, samples of intact or solubilized virus (dissociated with 0.25% Triton X-100 for 15 min at 23°C), were diluted 1:1 with 0.2 M triethanolamine, pH 8.5, and cross-linked at 23°C for 15 min following the addition of 0.02 volumes of the DTSP stock solution. DMS cross-linked samples were prepared by the addition of an equal volume of the DMS stock solution and incubated at 23°C for 1 hr. Noncross-linked control samples were treated identically except for the omission of the cross-linking reagent from the stock solutions. If samples were used only for analytical electrophoresis, cross-linking was terminated by the addition of sodium dodecylsulfate (Pierce). Alternatively, samples to be immuno-

precipitated were incubated for 15 min at 23°C with at least a ten-fold molar excess of lysine (pH 8.5) over the amount of the cross-linking reagent. Cross-linker concentrations are given as the final concentration in the reaction mixture.

Infection and Labeling of Monolayers. Confluent CEF monolayers in 60 mm petri plates were infected at high multiplicity (20-50 plaque forming units/cell) with Sindbis virus or mock infected at 37°C in the presence of actinomycin D (0.5 μg/ml) as previously described (Pierce et al., 1974). For long periods of radiolabeling (>10 min), the inoculum was replaced with new medium containing 1/40 the normal concentration of leucine which during labeling contained 50 μCi/ml L-^{3}H-leucine (Amersham, specific activity 52 Ci/mmole). In pulse-chase experiments the inoculum was replaced with regular medium at 1 hr post infection. Immediately before the pulse, the monolayers were washed twice with prewarmed medium lacking methionine, and pulse-labeled at 37°C for 10 min using medium lacking methionine containing 20 μCi/ml L-^{35}S-methionine (Amersham, specific activity 1075 Ci/mmole). Following the pulse, the plates were washed twice with prewarmed medium containing twice the normal concentration of methionine and incubated with 5 ml of this medium for the chase periods indicated.

Prelabeled monolayers were prepared by labeling ~1/3 confluent CEF monolayers with regular medium containing 1/4 the normal concentration of leucine and 50 μCi/ml ^{3}H leucine. At confluency, the monolayers were washed 3 times with regular medium containing 5 times the normal leucine concentration over a period of 1.5 hr. Monolayers were infected or mock infected and incubated in the presence of this medium.

All monolayers were prepared for subsequent cross-linking or solubilization by chilling on ice followed by at least 3 washes of ice cold phosphate buffered saline (PBS) (Dulbecco and Vogt, 1954, lacking Ca^{++} and Mg^{++}). Triton X-100 solubilized monolayers were lysed on ice in 1 ml of 0.5% Triton X-100, 0.05 M triethanolamine, pH 8.5, 0.2 M NaCl and nuclei were removed by centrifugation at 900 x g for 5 min

at 4°C. Control and DTSP cross-linked samples of these supernatants were prepared as described above for cross-linking of virions.

Cross-linking of Intact CEF Monolayers. Stock solutions of DTSP dissolved in DMSO were diluted 1/50 into ice cold PBS immediately before use. Monolayers were cross-linked with 2 ml of the PBS-DTSP solution (per 60 mm petri plate) by incubation on ice for 10 min. Noncross-linked controls were incubated in the presence of 2% DMSO in PBS. Following three washes of 25 mM Tris-Cl, 50 mM lysine, 0.1 M NaCl, pH 7.4, over the course of 10 min on ice, the monolayers we collected by solubilization in 1 ml of 0.5% SDS, containing 20 μg/ml phenylmethylsulfonylfluoride at 23°C. DNA was sheared by repeated pipetting of the lysate and the samples stored at -70°C.

Preparation of Antisera. Sindbis virus was purified and structural proteins were prepared as previously described (Bell et al., 1979). The purified proteins were precipitated overnight with 2.5 volumes of 100% EtOH at -20°C, warmed to 4°C, and centrifuged at 15,000 r.p.m. for 15 min at 4°C in a Sorvall SS34 rotor. The supernatants were discarded and the pellets air-dried and resuspended in 10 mM Tris pH 7.4, 0.15 M NaCl, 0.1% SDS to a final protein concentration between 100 and 200 μg/ml. The antigens were emulsified with an equal volume of complete Freund's adjuvant (Calbiochem) immediately before injection. Female New Zealand White rabbits were given multiple subcutaneous and intradermal injections along the lower back, as well as in intramuscular injection in each hind leg near the lymph node. 100-200 μg of each protein were used for the primary immunization and 25-100 μg for subsequent challenges. Immune and non-immune animals were bled from the ear vein and clot formation allowed to proceed for at least 1 hr at 23°C. After incubation for at least 24 hr at 4°C the clots were removed by centrifugation at 5000 x g for 15 min at 4°C. Serum was stored frozen at -70°C. Crude gamma globulin fractions were prepared by two sequential ammonium sulfate fractionations as described by Garvey et al. (1977). For preparation of IgG, the crude gamma globulin fraction was dialyzed exhaustively

against 10 mM PO_4, pH 7.2, 15 mM NaCl. After centrifugation at 100,000 x g for 20 min at 4°C, the supernatant was passed over a column of CM52 (Whatman), then a column of DE52 (Whatman) both equilibrated in 10 mM PO_4, pH 7.2, 15 mM NaCl (Palacios et al., 1972). The excluded IgG containing fractions were collected, pooled, and stored frozen at -70°C.

Hybrid clones of BALB/c MOPC21 NSI/1 (courtesy of C. Milstein) and spleen cells from BALB/c mice immunized with Triton X-100 solubilized Sindbis virus (1 mg Triton X-100/mg virus protein) were produced essentially as described by Nowinski et al. (1979). The production and characterization of these clones will be described in detail elsewhere. One of these clones, 6-8G, was found to secrete IgG specific for E1. IgG was purified from the culture medium by passage over a column of Protein A-Sepharose CL-4B (Pharmacia Fine Chem. Co.) and eluted with 0.1 M glycine, pH 3.0. After dialysis against PBS, the IgG was stored frozen at -70°C in small aliquots.

Immunoprecipitation. Samples were adjusted to 50 mM Tris pH 7.4, 0.2 M NaCl, 1 mM EDTA, 0.5% Triton X-100 and 1 mg/ml BSA (TNA buffer) by dilution. In some cases the concentrations of Triton X-100 and BSA were lowered to 0.1% and 200 µg/ml, respectively, in order to facilitate electrophoretic analysis of the unprecipitated supernatant. SDS-containing samples were heated to 56°C for 10 min to dissociate aggregates prior to dilution into this buffer. After dilution, there was at least a five-fold excess of Triton X-100 over SDS by weight. Immunoprecipitation of Triton X-100 solubilized monolayers was done immediately after solubilization without freezing. Rabbit IgG or monoclonal 6-8G anti-E1 IgG were diluted into TNA. Incubation with the antibody was for 30 min at 23°C, followed by removal of the immune complexes by a 10 min incubation with an excess of TNA washed protein A-bearing Cowan I strain of Staphylococcus aureus (Kessler, 1975) and centrifugation at 3500 x g for 6 min. Immunoprecipitates were washed with 50 mM Tris pH 7.4, 0.2 M NaCl, 1 mM EDTA, 0.1% Triton X-100, 200 µg/ml BSA.

Gel Electrophoresis. For the resolution of cross-linked species a continuous gel system similar to that of **Davies and Stark (1970)** was used. Gels contained 4% acrylamide (Bio Rad), 0.11% bisacrylamide (Bio Rad), 0.05 M sodium acetate, 0.05 M borate, 0.1% SDS, pH 8.5. Samples were dissociated with an excess of SDS, with or without 20 mM dithiothreitol (DTT) (Sigma), and heated to 56°C for 30 min prior to electrophoresis.

Analytical cylindrical gels to be quantitated were frozen in dry ice, sliced in 1 mm fractions using a Mickle gel slicer, and counted after shaking for at least 24 hr in 10 ml of a scintillation fluid composed of toluene, Liquifluor (NEN), and NCS (Amersham). Preparative cylindrical gels were sliced and eluted into siliconized glass vials containing 0.02% SDS, 2 mM DTT, 20 μg/ml phenylmethyl sulfonylfluoride 10 μl samples of each fraction were counted to determine the positions of cross-linked species and the peak fractions were pooled and stored frozen at -70°C.

Discontinuous electrophoresis was performed essentially as described by Laemmli (1970) except that the Tris buffer concentration in both the stacking and separating gels was halved. Bromophenol blue was used as the tracking dye in both gel systems. Details of the sample preparation will be presented in each experiment.

Two-dimensional Electrophoresis. Continuous gels in 4 mm id. tubes were used for the first dimension. After electrophoresis, the gel was removed from the tube and equilibrated for 30 min in 50 mM Tris pH 6.8, 1% β-mercaptoethanol, 40 mM DTT, and 1% SDS with continuous shaking. The gel was anchored to the second dimension using a hot solution of 1% agarose (Seakem), 50 mM Tris pH 6.8, 1% β-mercaptoethanol, 10 mM DTT, and 0.1% SDS. The second dimension was the discontinuous system described above using a 1.5 mm thick, 10% acrylamide separating gel, and a 4% acrylamide, 3 cm long, stacking gel. Electrophoresis was performed with 0.025% 3-mercapto-proprionic acid present in the top electrode buffer reservoir to prevent possible reoxidation of reduced disulfides during electrophoresis.

All slab gels were fixed in 10% acetic acid, 25% methanol. After treatment

for fluorography according to the method of Bonner and Laskey (1974), they were dried and exposed at -70°C using pre-fogged (Laskey and Mills, 1976) Kodak X-Omat R film.

RESULTS

Cross-linking of Sindbis Virus and SFV. We will briefly summarize our data from preliminary cross-linking studies of Sindbis virus and SFV with respect to several variables: type of cross-linker and concentration dependence, temperature of cross-linking, method of virus isolation, and cell type of virus growth.

Both DMS and DTSP (see below), as well as dimethyl 3,3' dithiobis proprionimidate (data not shown), gave similar patterns with each virus as would be expected since they react primarily with lysine and are capable of establishing cross-links within a distance of about 11 Å. Prior treatment of samples with SDS abolished cross-linking. The concentration of virus protein (ranging from 20 µg/ml to 1 mg/ml) or the inclusion of carrier BSA (100 µg/ml) did not effect the cross-linking patterns. Identical results were obtained with Sindbis virus isolated by several different methods: the large scale preparative method of Bell et al. (1979), the salt-reversal method (Pierce et al., 1974) or direct isolation of the virus from the culture medium by a single isopycnic centrifugation. In addition, isolation of Sindbis virus or SFV by successive cycles of potassium tartarate isopycnic centrifugation and pelleting (Kääriänen et al., 1969) did not alter their cross-linking patterns. Cross-linking of Sindbis virus at 0°C, 23°C, or 33°C gave similar patterns of cross-linked species, and although at lower temperatures the rate of cross-linking by DMS was much slower, DTSP cross-linking was rapid even at 0°C (Lomant and Fairbanks, 1976). In addition, cross-linking patterns of either Sindbis virus of SFV grown in CEF or BHK-21 cells were the same.

As can be seen in Figs. 1 and 2, cross-linking of Sindbis virus or SFV with DMS or DTSP produced relatively complex patterns of cross-linked species. Identification of the cross-linked glycoprotein oligomers was simplified by removal of the capsid

Fig. 1. DMS cross-linking of CEF- and BHK-grown Sindbis virus. CEF-grown (lanes 1, 2, 4, 5 and 6) or BHK-grown (lanes 3 and 7) Sindbis virus was cross-linked with 0.5 mg/ml DMS. Samples in lanes 2, 3, 6 and 7 were solubilized with Triton X-100 prior to cross-linking. After quenching with lysine, and dilution into TNA, the cross-linked capsid protein species were removed by immunoprecipitation with rabbit αC IgG as described in the Materials and Methods. The supernatants, containing the cross-linked glycoprotein species, were precipitated with 2.5 volumes of absolute ethanol. Greater than 99% of the counts were precipitated. Samples were resuspended in 1% SDS, 10% glycerol, 40 mM DTT, heated to 56°C for 30 min, and a portion of each run on a 4% acrylamide continuous slab gel. Lanes 1-3 are the rabbit αC immunoprecipitates, and lanes 5-7 the immunoprecipitation supernatants. Lane 4 is cross-linked intact virus without immunoprecipitation. In this figure as well as in subsequent figures, the cross-linked glycoprotein and capsid protein multimers are labeled G1, G2, . . ., and C1, C2, . . ., respectively.

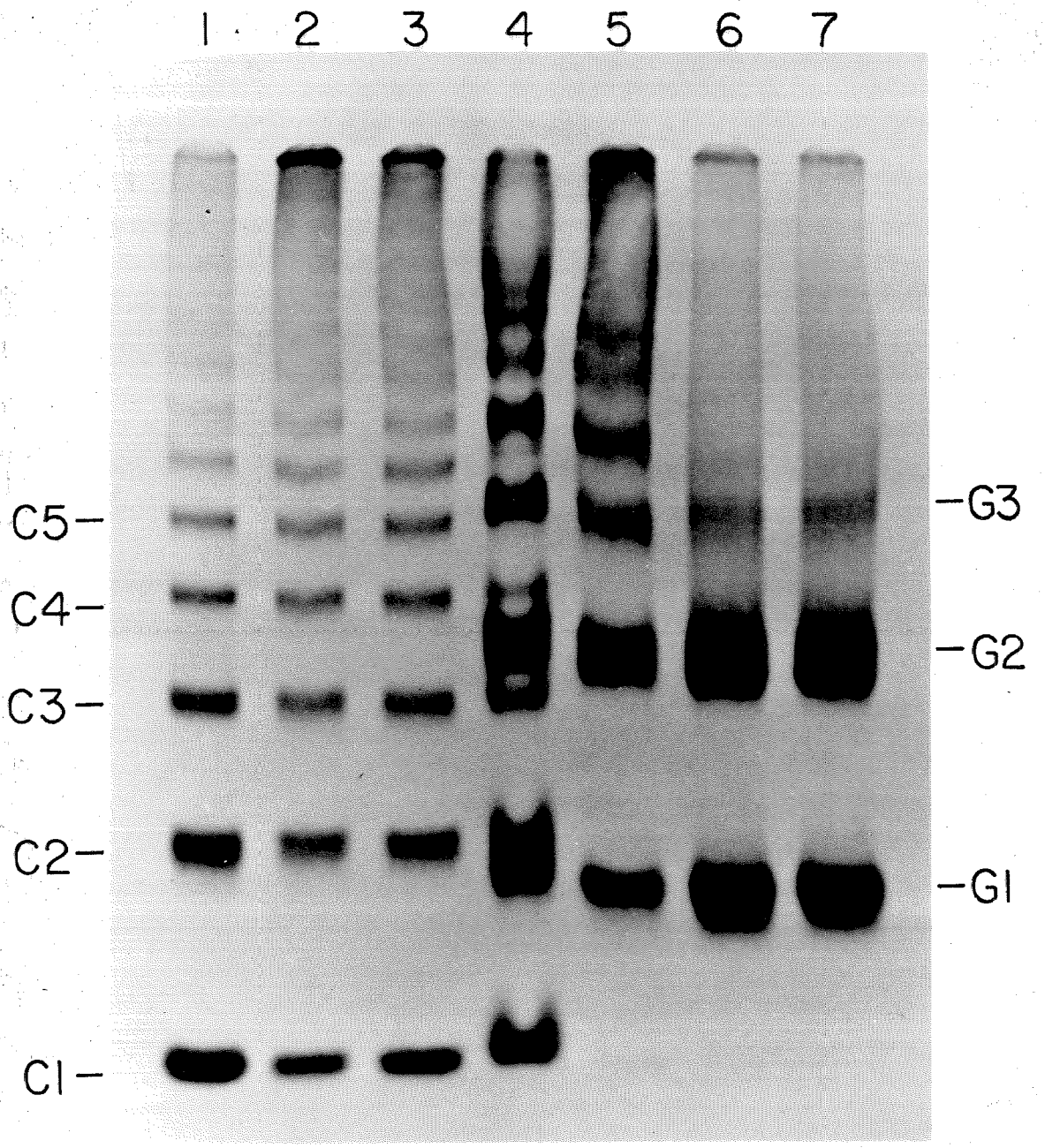
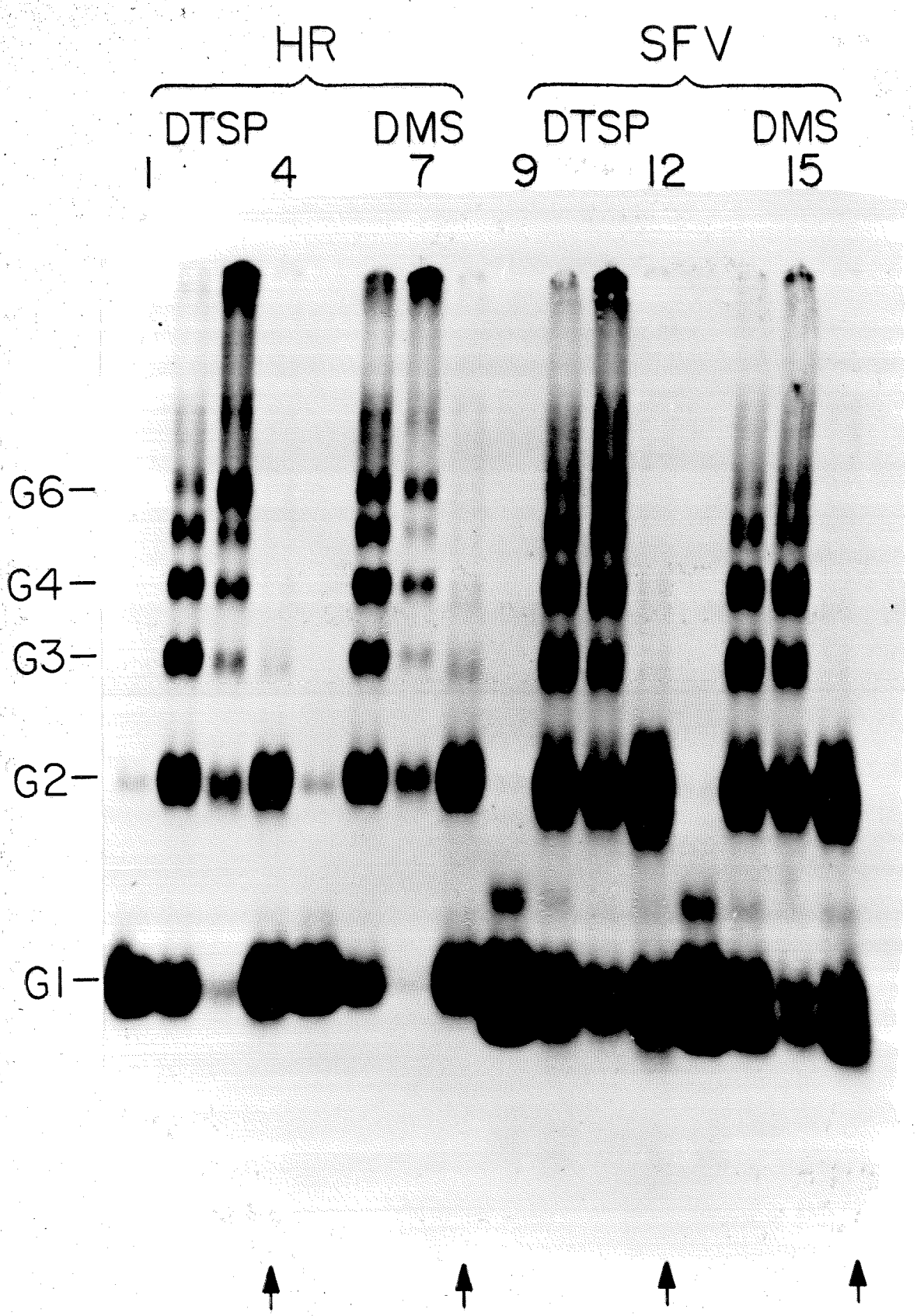

1
2
3
4
5
6
7
C5
C4
C3
C2
C1
G3
G2
G1

Fig. 2. Cross-linking of Sindbis virus and SFV with DMS and DTSP. Samples of BHK-grown Sindbis virus (HR strain) (lanes 1-8) or SFV (lanes 9-16) were cross-linked with DMS or DTSP as described in the Materials and Methods. Samples in lanes 4, 8, 12 and 16 were solubilized with 0.25% Triton X-100 prior to cross-linking (indicated by arrows). DTSP concentrations were 0.02 mg/ml (lanes 2 and 4), 0.06 mg/ml (lane 3), 0.1 mg/ml (lanes 10 and 12), and 0.2 mg/ml (lane 11). DMS concentrations were 0.5 mg/ml (lanes 6 and 8), 1.0 mg/ml (lanes 14 and 16), 1.5 mg/ml (lane 7), and 3 mg/ml (lane 15). Samples in lanes 1, 5, 9 and 13 were noncross-linked controls. After cross-linking, each sample was quenched with lysine and solubilized with Triton X-100. The nucleocapsids were removed by immunoprecipitation with rabbit αC IgG. The unprecipitated supernatants, containing the cross-linked glycoprotein species, were diluted with an equal volume of 4% SDS, 20% glycerol, heated to 56°C for 30 min and run on a 4% acrylamide slab gel.


HR
SFV
DTSP
DMS
DTSP
DMS
1
4
7
9
12
15
G6–
G4–
G3–
G2–
G1–

protein and its oligomers using a rabbit antiserum to the Sindbis virus capsid protein. This antiserum also reacted with the SFV capsid protein (Fig. 2) (Dalrymple et al., 1976). Prior to immunoprecipitation, the cross-linking reactions were quenched by the addition of excess lysine and dissociated with Triton X-100. In addition to the complete removal of the cross-linked capsid protein species (Figs. 1 and 2), no detectable glycoprotein aggregates were visible in the immunoprecipitates from either intact or Triton X-100 solubilized virions cross-linked with DMS (Fig. 1, lanes 1-3). Cross-linking of intact virions resulted in the formation of both glycoprotein multimers and capsid protein multimers. Raising the cross-linker concentration favored the formation of higher-order multimers, and in the case of the envelope glycoproteins these high concentrations produced a pattern in which even multimers were more predominant than the next lowest odd multimer (i.e., tetramer more predominant than trimer) (Fig. 2, lanes 3, 7, 11 and 15) as has been reported for SFV (Garoff, 1974). At lower cross-linker concentrations, however, even glycoprotein multimers were not preferentially formed over the next lowest odd multimer (Fig. 2, lanes 2, 6, 10 and 14). The pattern of discrete cross-linked glycoprotein oligomers terminates abruptly with the hexamer, although at high cross-linker concentrations larger aggregates barely entering the gel were produced. When the virions were disrupted with an excess of Triton X-100 followed by cross-linking, the glycoproteins were cross-linked primarily into dimers with minor amounts (less than 5%) of trimers and tetramers (Fig. 1; Fig. 2, lanes 4, 8, 12 and 16). Apparently, a stable association between the Sindbis virion glycoproteins persists after Triton X-100 solubilization as has been found for SFV (Garoff, 1974).

Immunoprecipitation of DMS Cross-linked Glycoprotein Dimers. We wished to determine whether the glycoprotein dimer band from either intact or Triton X-100 solubilized cross-linked virions was a pure population of glycoprotein homodimers, heterodimers, or a mixture of both. In order to answer this question SDS-denatured

E1 and E2 were used to produce rabbit antisera specific for each virion glycoprotein. These antisera were specific as shown by the immunoprecipitates obtained with SDS-denatured Sindbis virus (Fig. 3).

The proportion of the cross-linked glycoprotein dimer band precipitable by an excess of each glycoprotein specific IgG or a combination of both antisera was used to determine the proportion of heterodimers. Nonimmune rabbit IgG was used as a control for nonspecific precipitation (always less than 2%). In all cases, greater than 95% of the input was precipitated by a combination of both glycoprotein antisera. Two different procedures were used. In the first method (shown in Table 1), the cross-linked dimer band was isolated from preparative gels and immunoprecipitated with glycoprotein specific IgG. The results show that the cross-linked dimer band is more than 90% heterodimers. To rule out the possibility that homodimers were selectively lost during the preparation of the cross-linked dimer band, we directly immunoprecipitated equal proportions of DMS cross-linked samples with each glycoprotein specific IgG or a combination of both. The immunoprecipitates were run on analytical gels, and the various cross-linked species quantitated (Fig. 4). For cross-linked intact or Triton X-100 solubilized Sindbis virus, both methods showed that more than 90% of the glycoprotein dimers were heterodimers of E1 and E2, whether the virus was grown in CEF or BHK-21 monolayers. The inability to precipitate all of the dimers with each antisera may reflect either the presence of a small proportion of homodimers, or the loss of reactivity of some of the heterodimers with one of the glycoprotein specific IgGs. Since the majority (>65%) of the Triton X-100 solubilized glycoproteins can be cross-linked into dimers using higher cross-linker concentrations (data not shown), the above results suggest that the E1-E2 heterodimer is the predominant form of the glycoproteins following Triton X-100 solubilization.

The specificities of the anti-glycoprotein IgGs were not directly demonstrable on the glycoprotein monomers in DMS-treated samples since these monomers of E1

Fig. 3. Specificity of antisera to Sindbis structural proteins. Rabbit αE1 IgG, αE2 IgG or αC IgG as well as mouse monoclonal αE1 IgG (clone 6-8G) were prepared as described in the Materials and Methods. SDS-denatured Sindbis virus (^{35}S-methionine labeled) (lanes 1, 6, 7) in TNA was precipitated with rabbit αC (lane 5), monoclonal αE1 (lane 8), or nonimmune rabbit IgG (lane 9). The unprecipitated supernatant from the αC immunoprecipitation was then precipitated with rabbit αE1 (lane 2), αE2 (lane 3), or nonimmune (lane 4) IgG. The washed immunoprecipitates were resuspended in sample buffer containing 20 mM DTT, and heated to 56°C for 30 min. After centrifugation at 3500 x g, for 6 min, the supernatants were run on a 10% Laemmli slab gel. A trace amount of PE2 can be seen in lanes 1, 3 and 6.

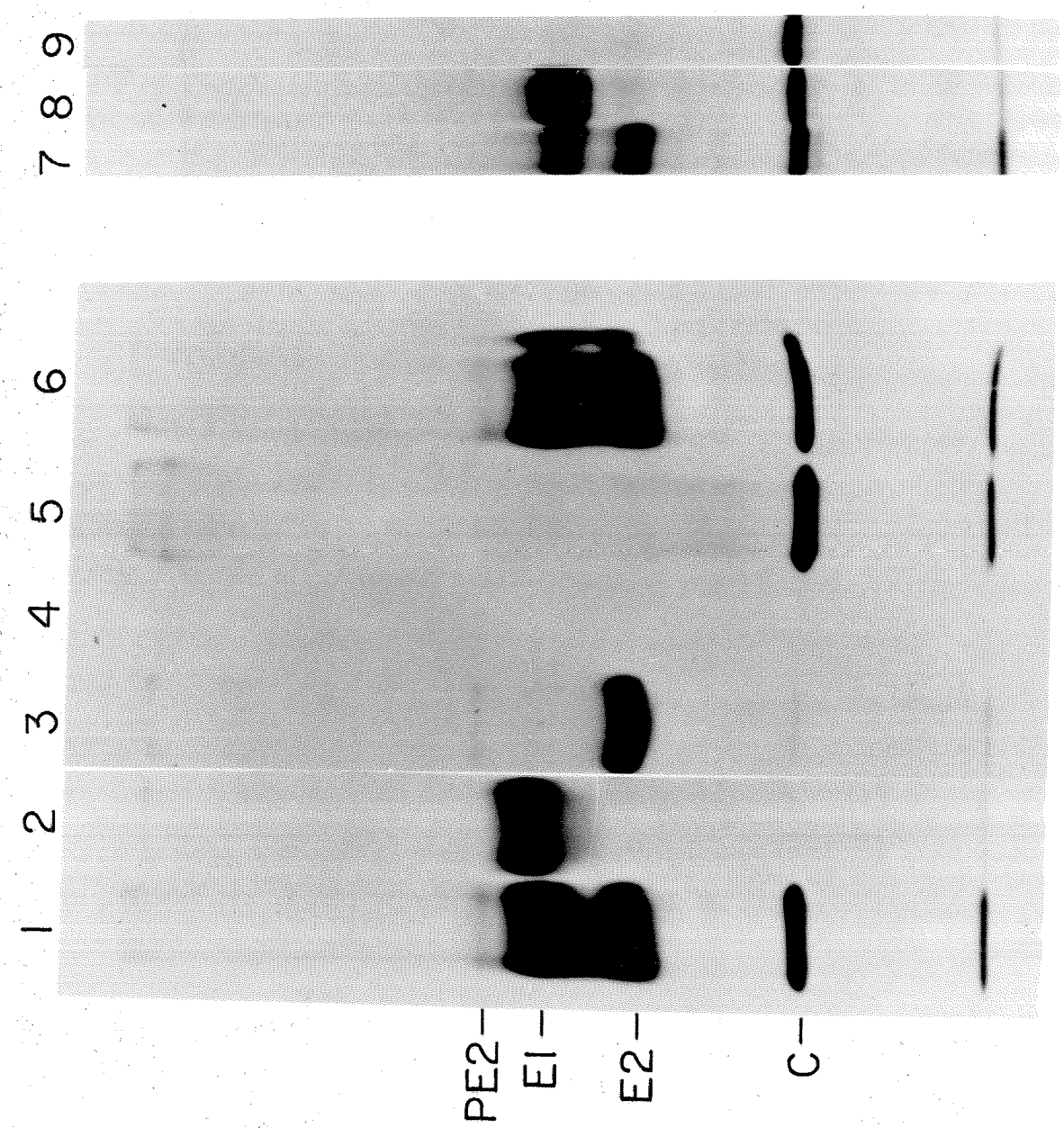

1
2
3
4
5
6
7
8
9
PE2
E1
E2
C

Table 1

Immunoprecipitation of DMS Cross-linked Sindbis Virus Glycoprotein Dimers[a]

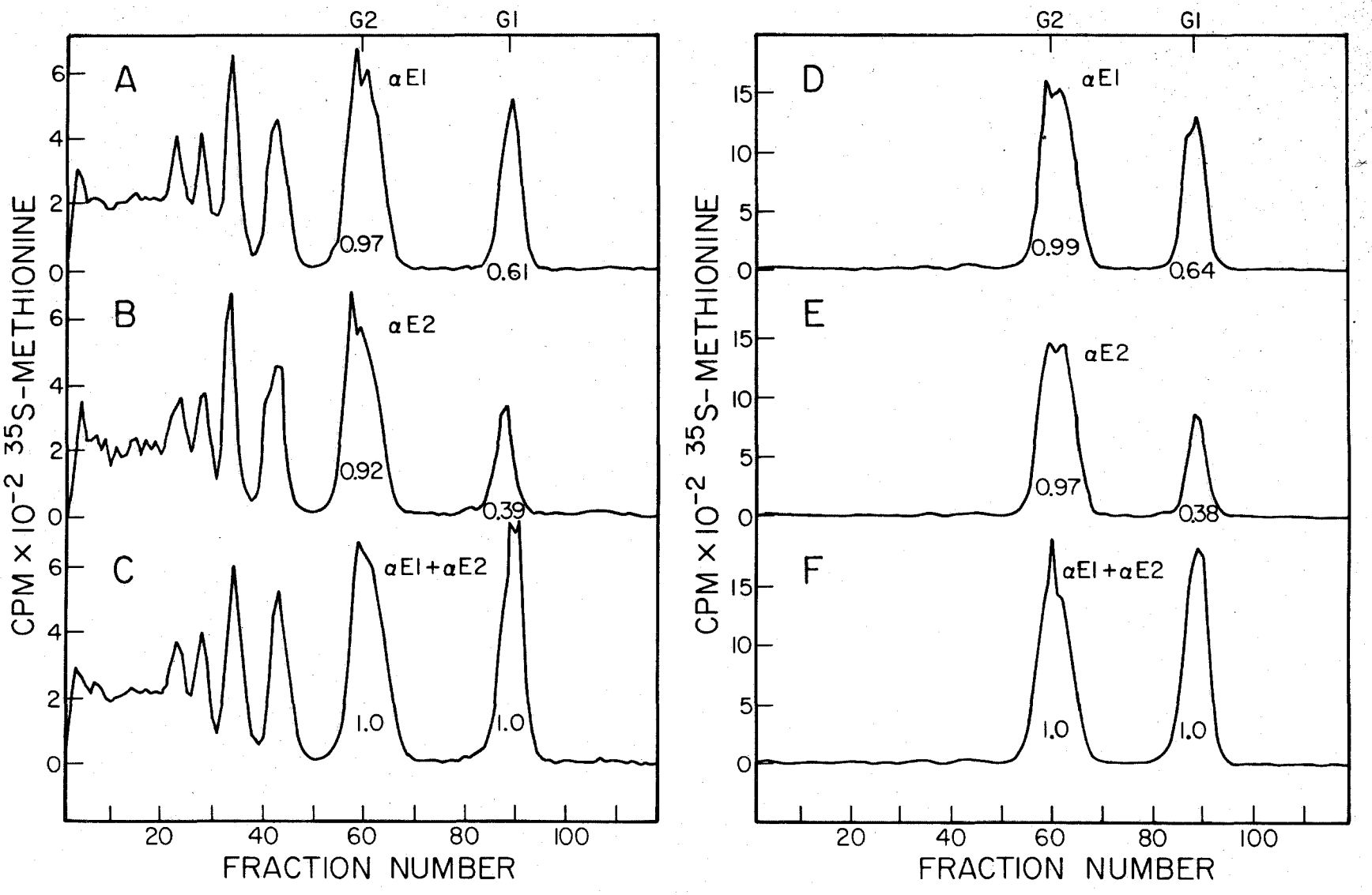
	IgG	% precipitated[b]	% of precipitable cpm[c]
CEF-grown Intact	αE1	94.4	95.0
	αE2	94.4	95.4
	αE1 + αE2	99.4	100.0
	nonimmune	1.6	0.0
CEF-grown Triton X-100 solubilized	αE1	90.7	94.9
	αE2	92.1	96.1
	αE1 + αE2	95.9	100.0
	nonimmune	1.2	0.0
BHK-grown Triton X-100 solubilized	αE1	94.3	94.8
	αE2	95.9	96.4
	αE1 + αE2	99.4	100.0
	nonimmune	1.3	0.0

[a] Samples prepared as described in Fig. 1 (lanes 5, 6 and 7) were run on preparative 4% acrylamide cylindrical gels. The gels were sliced and the cross-linked glycoprotein dimer band was isolated as described in the Materials and Methods.

[b] % Precipitated of the total cpm recovered. Recovery was >96% of the input in all cases.

[c] The cpm specifically precipitated by each antisera was normalized to the total cpm specifically precipitable by a combination of both antisera.

Fig. 4. Immunoprecipitation of DMS cross-linked Sindbis virus. Samples of intact (panels A, B and C) or Triton X-100 dissociated (panels D, E and F) Sindbis virus grown in BHK-21 cells were cross-linked with DMS as described in Fig. 1. After removal of the capsid protein (see Fig. 1), and denaturation with SDS, each sample was divided into three equal portions and immunoprecipitated with αE1 (panels A and D), αE2 (panels B and E), or a combination of both glycoprotein-specific antisera (panels C and F). A combination of both antisera precipitated greater than 98% of the input cpm. The immunoprecipitates were solubilized in 1% SDS and 20 mM DTT and analyzed on 4% acrylamide cylindrical gels as described in the Materials and Methods. Electrophoresis was from left to right. The number under each peak represents the proportion of that glycoprotein oligomer which is precipitable by each antisera as defined in Table 1, footnote c.


INTACT SINDBIS
TRITON X-100 SOLUBILIZED SINDBIS
G2
G1
A
αE1
0.97
0.61
B
αE2
0.92
0.39
C
αE1+αE2
1.0
1.0
D
αE1
0.99
0.64
E
αE2
0.97
0.38
F
αE1+αE2
1.0
1.0
CPM × 10-2 35S-METHIONINE
FRACTION NUMBER
0
2
4
6
5
10
15
20
40
60
80
100

and E2 were no longer separable on 10% Laemmli gels (data not shown), probably due to intra-molecular cross-linking. However, one would expect that if aggregation of cross-linked species or cross-reaction of the antisera occurred with DMS cross-linked samples, then the sum of the glycoprotein monomer counts precipitable by each antisera should exceed the total number of monomer counts precipitable by both antisera (as it does for the dimers). However, in both methods, immunoprecipitation of the monomer band gave additive results (see Fig. 4).

We also examined the immunoprecipitates obtained with the glycoprotein-specific antisera using Triton X-100 solubilized virus which had not been cross-linked (data not shown). Antiserum to E2 precipitated a small (less than 10%) but certainly not equimolar amount of E1. Since the Triton X-100 solubilized glycoproteins were predominantly in the form of E1-E2 heterodimers, this result indicates that the interaction between E1 and E2 is relatively unstable and can be disrupted by incubation with glycoprotein-specific antisera.

Cross-linking of Infected Monolayers. In order to analyze the association of E1 and E2 in infected cells, the mercaptan cleavable cross-linking reagent DTSP was used to cross-link CEF monolayers infected with Sindbis virus (HR strain). Cross-linking of monolayers was carried out at 0°C to minimize artifacts due to diffusion limited collisions between proteins in the lipid bilayer (Birdwell and Strauss, 1974). The cross-linking reaction was quenched at 0°C by incubation with an excess of lysine before solubilization with SDS at 23°C. Samples were analyzed by two-dimensional diagonal electrophoresis. The first dimension separated cross-linked aggregates by size on a low percentage acrylamide gel under non-reducing conditions (Davies and Stark, 1979). After reduction, the polypeptide subunits of the disulfide-linked aggregates were resolved by electrophoresis on SDS polyacrylamide slab gels (Laemmli, 1970). Cross-linked species appear below the diagonal (see Fig. 5).

Fig. 5. Cross-linking of Sindbis virus (HR strain) infected monolayers early and late in infection. CEF monolayers were infected and labeled with ^{3}H-leucine as described in the Materials and Methods. After cross-linking with DTSP (0.2 mg/ml), and incubation with excess lysine, the monolayers were lysed in SDS and analyzed by two-dimensional electrophoresis. Labeling in panel A was from 2-3 hr postinfection, and in panel B from 2-6 hr postinfection. The first dimension was run from left to right, under nonreducing conditions, and the second dimension run from top to bottom, under reducing conditions. Cross-linked oligomers are indicated on the horizontal axes (as defined in Fig. 1) and the positions of the virion structural proteins and PE2 are shown on the vertical axis. All subsequent two-dimensional gels are presented in this format.

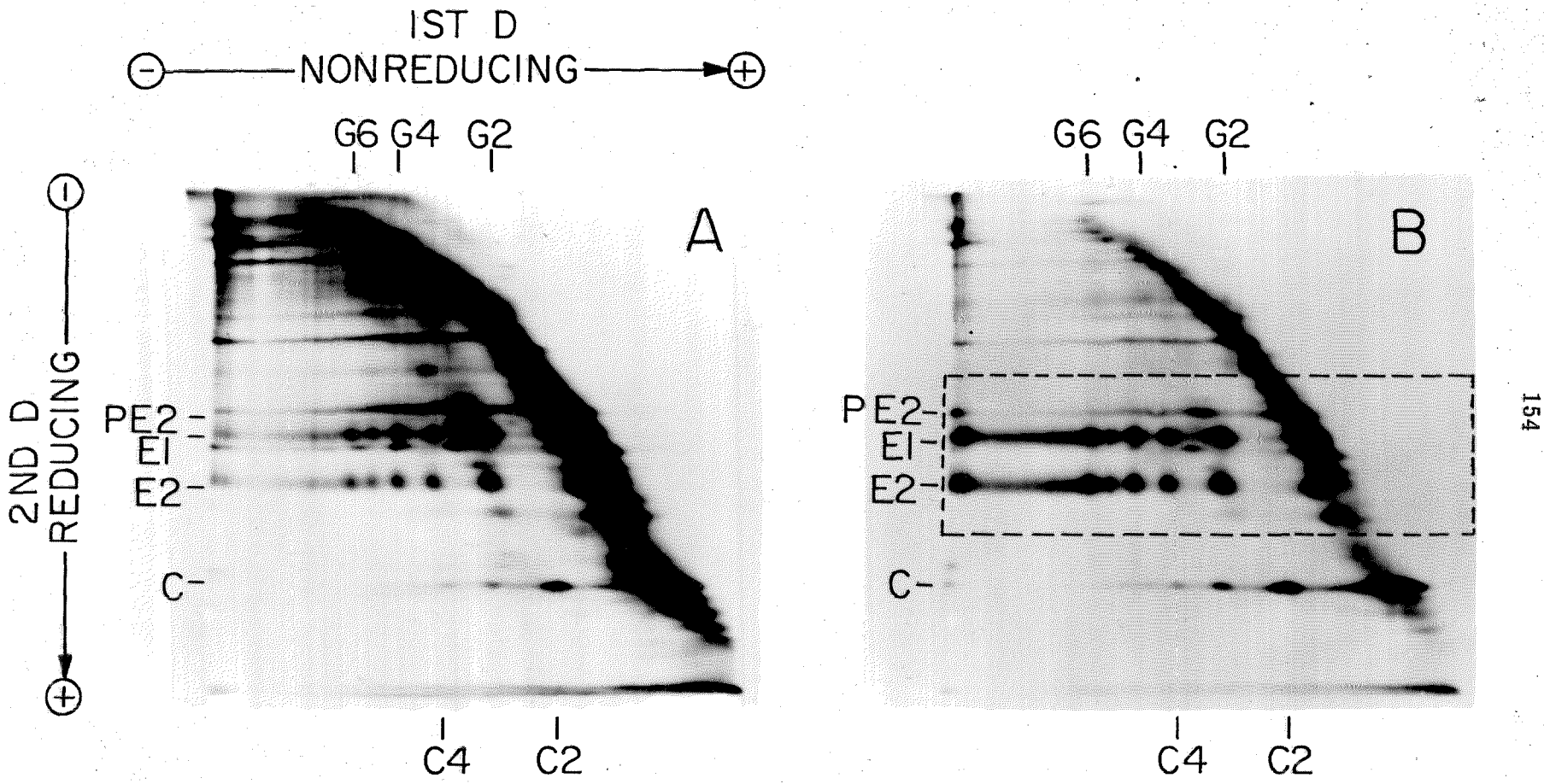

1ST D
NONREDUCING
G6 G4 G2
2ND D
REDUCING
PE2
E1
E2
C
A
C4 C2
G6 G4 G2
PE2
E1
E2
C
B
C4 C2

Figure 5 shows the two-dimensional patterns of Sindbis virus (HR strain) infected monolayers labeled from 2-3 hr postinfection (Fig. 5, panel A) and 2-6 hr postinfection (Fig. 5, panel B) and cross-linked with DTSP (0.2 mg/ml). By 3 hr postinfection, the glycoproteins have appeared in the plasma membrane, but substantial budding does not occur until 3.5-4 hr after infection (Birdwell and Strauss, 1974). The two-dimensional patterns of these cross-linked monolayers were qualitatively similar. A large number of noncross-linked and cross-linked polypeptides were found in the 2-3 hr labeling period (Fig. 5, panel A). These were mainly host cell proteins whose synthesis had not yet been shut off (data not shown). The predominant cross-linked host cell protein in mock infected as well as infected cells was a protein migrating slightly faster than E1 (see Fig. 5, panel A). In addition some higher molecular weight (>150,000 daltons) cross-linked polypeptides can be seen. E1 and E2 were cross-linked into regular multimers resolved clearly up to hexamers, while the capsid protein was less readily cross-linked. Monolayers incubated in the presence of 2% DMSO in PBS as a control did not contain any of these aggregates (data not shown; see Fig. 7). At this cross-linker concentration, there is clearly a predominance of cross-linked tetramers and hexamers of E1 and E2 over trimers and pentamers, respectively. It is of interest to note that although little budding was taking place at 3 hr after infection as compared with 6 hr after infection, the E1 and E2 cross-linking patterns were qualitatively similar, except for the presence of proportionately larger amounts of higher glycoprotein aggregates and a large amount of material hardly entering the first dimension at 6 hr after infection. Although PE2 was less readily cross-linked than either E1 or E2, cross-linked aggregates of PE2 were found in the dimer region and throughout the first dimension. In addition, at 6 hr after infection larger unresolved PE2 aggregates were found at the top of the first dimension.

These patterns of cross-linked E1 and E2 in infected cells suggested that they might be cross-linked preferentially to each other. However, since monomeric E1

and E2 were not separated in the first dimension, E1-E2 heterodimers could not be distinguished from viral glycoprotein homodimers or heterodimers of viral glycoproteins with unlabeled host cell proteins. In the case of PE2, the situation was even less clear since cross-linked PE2 aggregates were smeared throughout the first dimension, and even in the dimer region did not align with any other virus-specific proteins. We attempted to clarify the association by immunoprecipitation of the lysates with glycoprotein specific IgG followed by two-dimensional analysis. Fig. 6 shows the results of such an experiment, using a similar lysate to the one shown in Fig. 5, panel B. Precipitation with an antiserum to either E1 or E2 (Fig. 6) indicated that E1 and E2 were associated as heterodimers in the infected cell. The higher cross-linked aggregates were probably due to cross-linking of E1 and E2 to each other, rather than to unlabeled host cell protein since the two-dimensional patterns of cross-linked prelabeled host cell proteins (prelabeled during 2-3 cell divisions) were identical regardless of whether or not the cells had been infected (for 6 hr) with Sindbis virus (data not shown). The majority of cross-linked PE2 which could immunoprecitated with antiserum to E1 was present in large aggregates with very little material in the dimer region (Fig. 6, panel A). Additional data on cross-linked PE2 will be presented in subsequent sections.

Association of PE2 and E1 in Infected Cells. Our findings that the Sindbis virion envelope glycoproteins were associated as heterodimers, as well as similar results in the case of SFV (Ziemiecki and Garoff, 1978; C. M. Rice, unpublished), and that E1 and E2 were associated in infected cells suggested that a precursor PE2-E1 complex might also be present. The ability to detect such an association by cross-linking depends not only on the lability of the complex to cross-linking but also its steady-state levels in the infected cell. Previous experiments using infected monolayers labeled for relatively long periods of time failed to convincingly demonstrate the existence of PE2-E1 heterodimers (see Figs. 5 and 6). Therefore, we approached these problems by using pulse-chase experiments as well as by examining monolayers infected with temperature-sensitive mutants defective in the cleavage of PE2.

Fig. 6. Immunoprecipitation of cross-linked monolayers. A lysate prepared as in Fig. 5, panel B, was heated, then diluted into TNA. After preclearing with αC IgG, the supernatant was divided into three equal proportions and immunoprecipitated with antisera (IgG fraction) to either E1, E2, or a combination of both glycoprotein-specific antisera. The washed immunoprecipitates were resuspended in 4% SDS, 20% glycerol and heated to 56°C for 30 min. After centrifugation at 3500 x g for 6 min, the supernatants were analyzed on two-dimensional gels. Only the region of the second dimension containing the virion glycoproteins and PE2 is shown (analogous to the region in Fig. 5, panel B, enclosed by the dotted line).

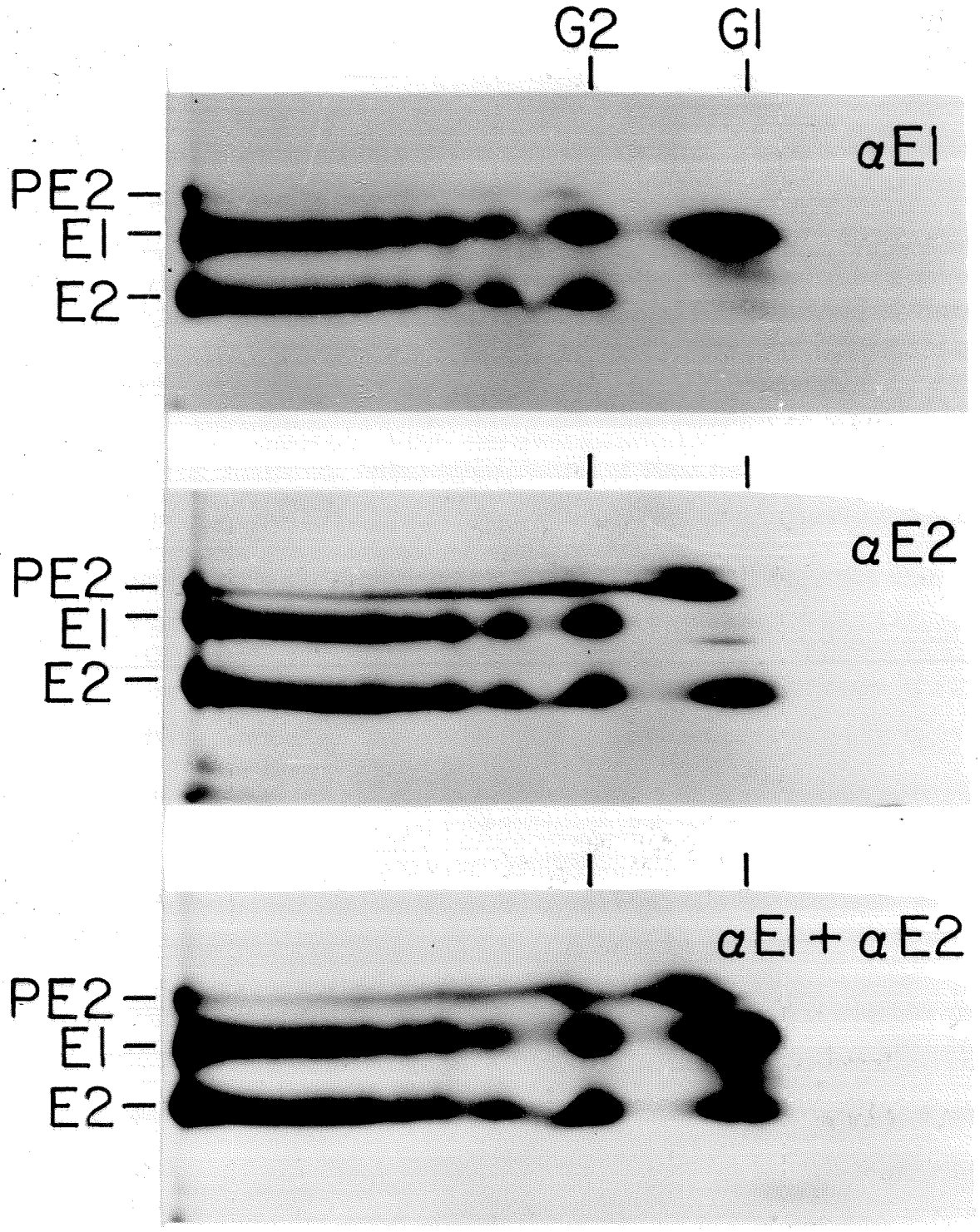

G2
G1
αE1
PE2
E1
E2
αE2
PE2
E1
E2
αE1 + αE2
PE2
E1
E2

Figure 7 shows the two-dimensional patterns of pulse-chased monolayers infected with Sindbis virus (HR strain). After the 10 min pulse E1 and PE2 were inefficiently cross-linked into a smear of aggregates with the majority of the label in the dimer region (Fig. 7, panel E). Cross-linked PE2 and E1 in the dimer region do not align exactly in the first dimension as would be expected for a pure population of heterodimers. In addition, even in the noncross-linked monolayers some aggregates of PE2 and E1 can be observed. During the chase, as PE2 was cleaved to E2, increasing amounts of E1 and E2 were present in discrete cross-linked species (see Fig. 7, panels G and H). With longer chase periods the amount of cross-linked PE2 decreased and the predominant cross-linked species shifted from the dimer region to an aggregate barely entering the first dimension. Immunoprecipitation of these lysates with rabbit antiserum specific for E1 and analysis under reducing conditions (Fig. 8) revealed that a small but significant amount of PE2 was cross-linked to E1 throughout the chase while the amount of cross-linked E2 increased up to 50 min after the pulse (Fig. 8, panel A). In noncross-linked monolayers (Fig. 8, panel B) a small and constant amount of PE2 also precipitated (see below). From these experiements, it was clear that a small amount of PE2 could be cross-linked to E1, but that E1 and E2 were much more readily cross-linked.

If a large amount of the PE1-E1 complex was present in infected cells, our inability to detect it could be explained by either a lack of readily cross-linkable groups between E1 and PE2, or if the complex could be cross-linked it might be inaccessible to the DTSP. Since E1-E2 heterodimers were stable to Triton X-100 solubilization, we tested the latter possibility by cross-linking pulse-labeled monolayers which had first been solubilized with this detergent. The cross-linked lysates were then immunoprecipitated with high-titer monoclonal antiserum against E1 (see Fig. 2 for specificity). The results of this experiment (shown in Fig. 9) demonstrated that heterodimers of PE2 and E1 could be detected after a 10 min pulse (Fig. 9, panel E), but after a 50 min

Fig. 7. Cross-linking of pulse-labeled monolayers. CEF monolayers were infected with Sindbis virus (HR strain), pulse-labeled for 10 min at 37°C with ^{35}S methionine, and chased as described in the Materials and Methods. After washing with ice-cold PBS, monolayers were cross-linked with 0.2 mg/ml DTSP (panels E-H) or incubated with 2% DMSO (panels A-D), treated with lysine, and lysed in SDS. Samples were adjusted to 2% SDS, 10% glycerol and analysed on two-dimensional gels. The low molecular weight products aligning with the virus glycoproteins in cross-linked and noncross-linked monolayers were artifacts due to proteolysis (data not shown).

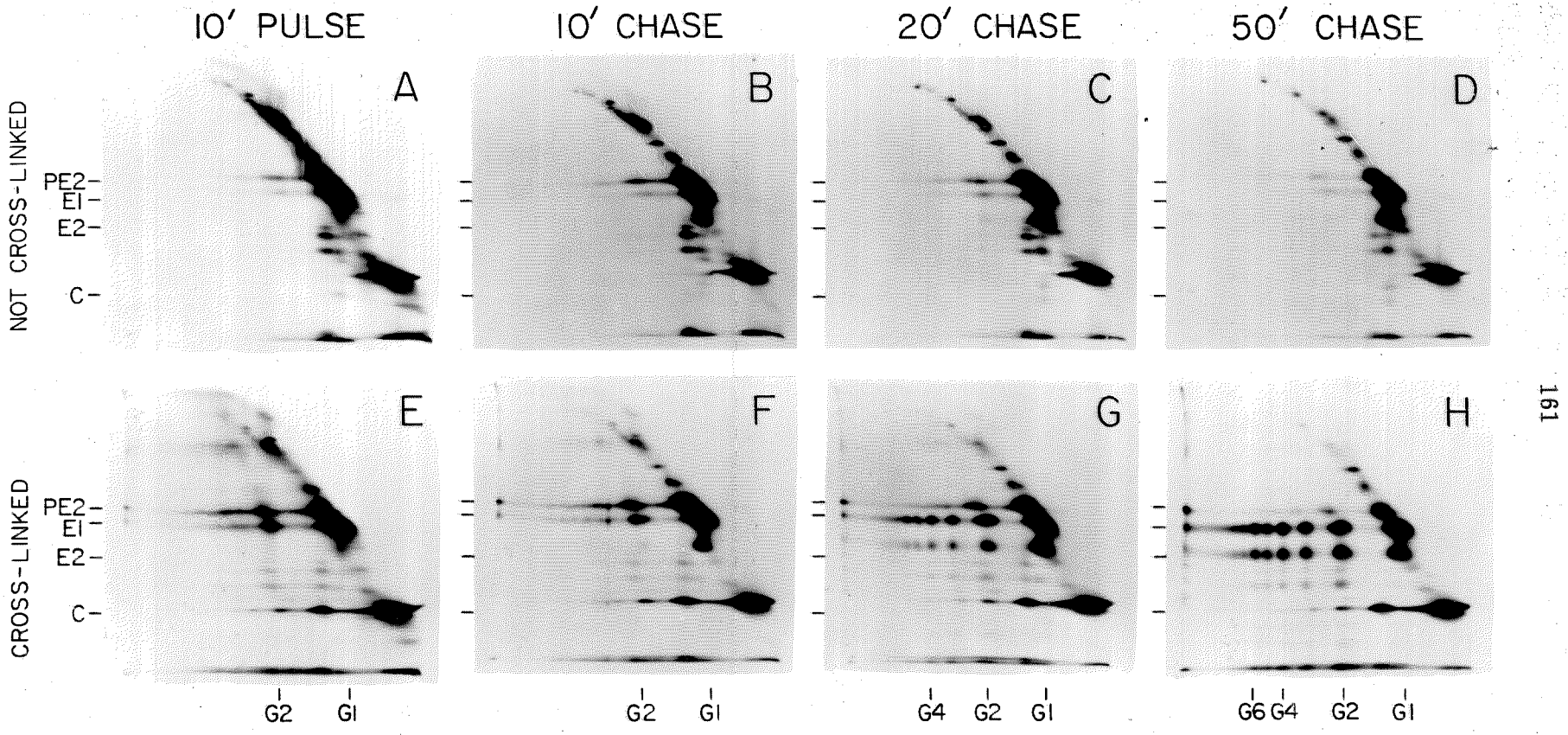

10′ PULSE
10′ CHASE
20′ CHASE
50′ CHASE
A
B
C
D
E
F
G
H
NOT CROSS-LINKED
CROSS-LINKED
PE2
E1
E2
C
G2
G1
G2
G1
G4
G2
G1
G6
G4
G2
G1

Fig. 8. Immunoprecipitation of cross-linked monolayers following a 10 min pulse and different times of chase with antiserum to E1. Samples were prepared as described in Fig. 7. A sample of each lysate was immunoprecipitated with rabbit αE1 IgG without prior preclearing with αC IgG. The washed immunoprecipitates and samples of each lysate were denatured with sample buffer containing 20 mM DTT, heated to 56°C for 30 min and run on 10% acrylamide slab gels (Laemmli, 1970). Cross-linked lysates are in panel A and noncross-linked lysates in panel B. Odd numbered lanes represent the lysate samples [indicated by (-)], and the corresponding αE1 immunoprecipitates [indicated by (+)] are in the next even lanes. Virus markers were run in the outside lanes.

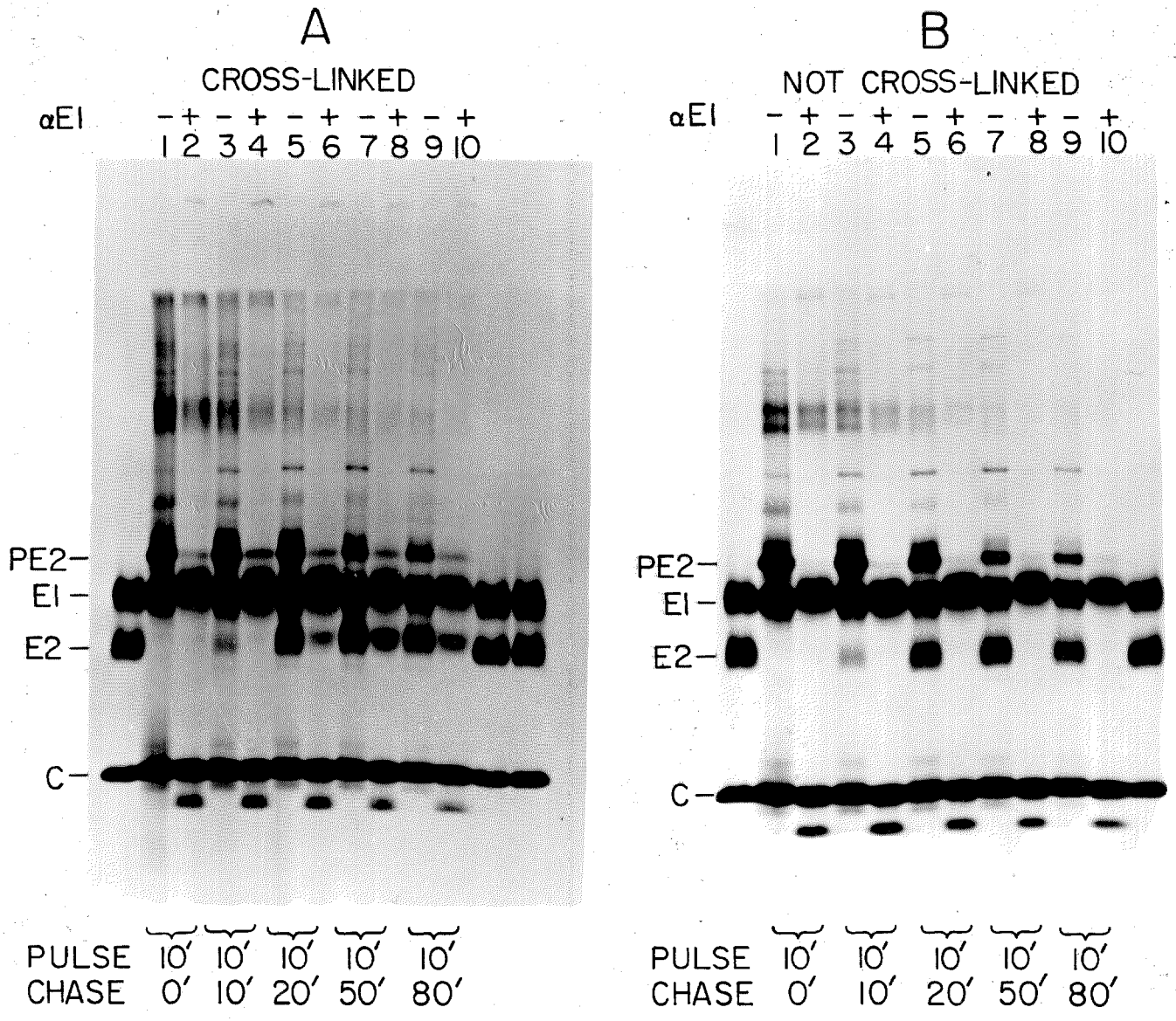
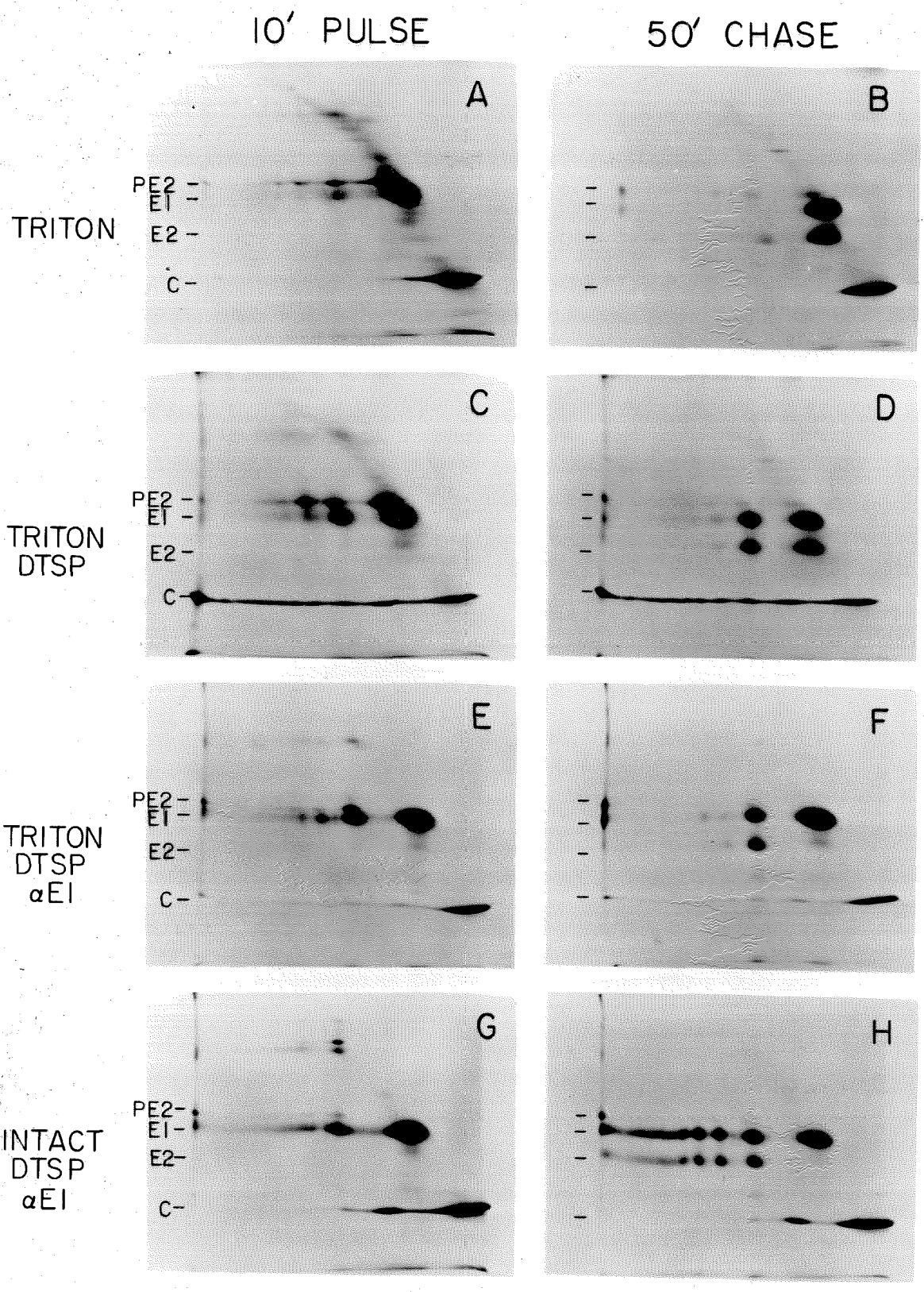

A
CROSS-LINKED
αE1
- + - + - + - + - +
1 2 3 4 5 6 7 8 9 10
PE2
E1
E2
C
PULSE 10′ 10′ 10′ 10′ 10′
CHASE 0′ 10′ 20′ 50′ 80′
B
NOT CROSS-LINKED
αE1
- + - + - + - + - +
1 2 3 4 5 6 7 8 9 10
PE2
E1
E2
C
PULSE 10′ 10′ 10′ 10′ 10′
CHASE 0′ 10′ 20′ 50′ 80′

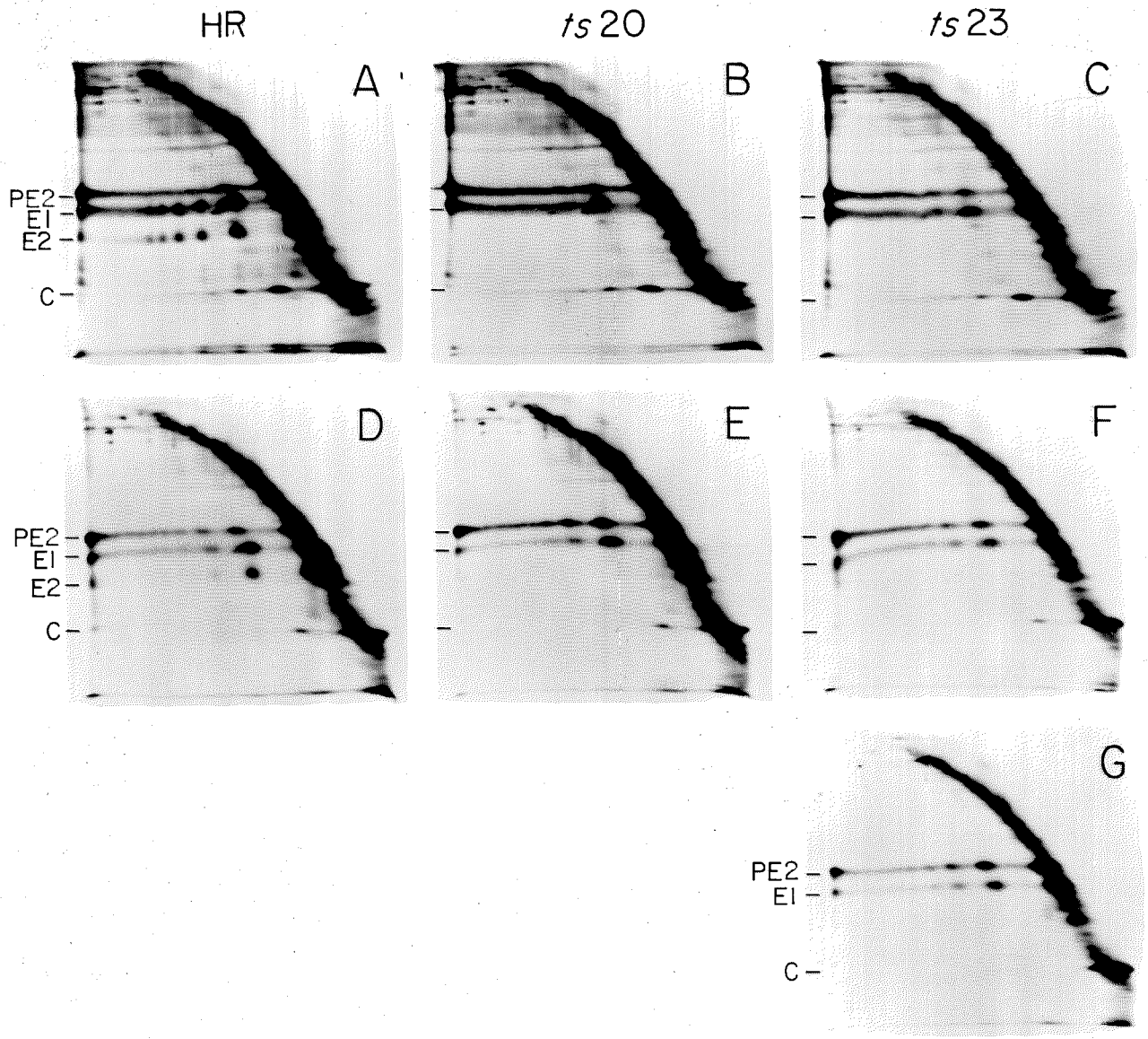
Fig. 9. Immunoprecipitation and two-dimensional analysis of infected monolayers cross-linked after Triton X-100 solubilization. Monolayers were infected and pulse-labeled only (panels A, C, E, and G), or pulse-labeled and chased for 50 min (panels B, D, F, and H), as described in Fig. 7. Monolayers in panels A and B were solubilized in Triton X-100 only; monolayers in panels C and D were solubilized with Triton X-100 then cross-linked with DTSP (0.04 mg/ml, room temperature, 15 min), panels E and F show the immunoprecipitates obtained with mouse monoclonal αE1 IgG of the material shown in panels C and D, respectively. The samples in panels G and H were intact monolayers, cross-linked with DTSP (as described in Fig. 7), lysed in SDS, and immunoprecipitated with monoclonal αE1 IgG. Before electrophoresis, all samples were denatured at 56°C for 30 min in 4% SDS, 10% glycerol, and analyzed on two-dimensional gels. The high-titer monoclonal αE1 IgG was used since it efficiently precipitated Triton X-100 or SDS solubilized E1 from lysates without overloading the gel systems used for further analysis. The faint spot running slightly faster than E1, precipitable by the monoclonal αE1 IgG, is probably an E1 degradation product rather than E2 contamination.


10′ PULSE
50′ CHASE
A
B
C
D
E
F
G
H
TRITON
TRITON DTSP
TRITON DTSP αEI
INTACT DTSP αEI
PE2
EI
E2
C

chase the PE2 cross-linked to E1 was present as larger aggregates. While solubilization of pulse-labeled monolayers prior to cross-linking did increase the recovery of PE2-E1 heterodimers (Fig. 9, compare panels E and G), the immunoprecipitation results (Fig. 9, compare panels C and E) suggested that the remainder of the cross-linked oligomers of E1 and PE2 did not involve PE2-E1 cross-linking. This conclusion is consistent with the observation that many of the cross-linked oligomers involving newly synthesized E1 and PE2 did not form discrete spots which aligned in the first dimension (Fig. 9, panels C, E and G; see also Fig. 7, panels A and E), as was found for E1-E2 oligomers following a 50 min chase (Fig. 9, panels D, F and H).

Since mutants from complementation groups D and E do not cleave PE2 at the nonpermissive temperature, we investigated the cross-linking patterns in cells infected with ts20 (group E) and ts23 (group D). Evidence has been presented that ts20 has a lesion in the gene coding for PE2 (Bracha and Schlesinger, 1976), contains viral glycoproteins in the plasma membrane (Bell and Waite, 1977; Smith and Brown, 1977; Saraste et al., 1980a) and allows nucleocapsid binding to the plasma membrane (Brown and Smith, 1975). ts23 contains a temperature-sensitive hemagglutinin (Yin, 1969), and the viral hemagglutinin has been identified as E1 (Dalrymple et al., 1976). Furthermore, nucleocapsids do not bind to the plasma membrane in cells infected with this mutant at the nonpermissive temperature (Brown and Smith, 1975). In addition, cells infected with this mutant contain little or no detectable viral glycoprotein (either E1 or E2) in the plasma membrane at the restrictive temperature (Bell and Waite, 1977; Smith and Brown, 1977; Saraste et al., 1980a). Figure 10 shows the two-dimensional patterns of Sindbis HR strain, ts20, and ts23 infected monolayers labeled at 40°C and cross-linked with 0.1 mg/ml DTSP or 0.02 mg/ml DTSP. The lower concentrations of DTSP were necessary in order to obtain significant quantities of PE2 and E1 oligomers entering the first dimension (see below). When compared to HR infected cells grown at 37°C (see Fig. 5, panel B), it is apparent that PE2 and E1 synthesized at 40°C were

Fig. 10. Cross-linking of monolayers infected with ts mutants. Confluent CEF monolayers were infected with Sindbis virus HR strain (panels A and D), ts20 (panels B and E), or ts23 (panels C, F and G) at 37°C. At 1.5 hr after infection monolayers were shifted to 40°C, and labeled with 50 μCi/ml ^{3}H leucine from 3.5-6 hr after infection. After washing with ice-cold PBS, the monolayers were cross-linked as described in the Materials and Methods with either 0.1 mg/ml DTSP (panels A-C), or 0.02 mg/ml DTSP (panels D-F), or incubated with 2% DMSO in PBS as a control (panel G). Following lysis with SDS, the samples were adjusted to 2% SDS, 10% glycerol, heated to 56°C for 30 min and analyzed on two-dimensional gels.


HR
ts 20
ts 23
A
B
C
D
E
F
G
PE2
E1
E2
C
PE2
E1
E2
C
PE2
E1
C

much more readily cross-linked in monolayers infected with HR as well as ts20 and ts23. The capsid protein cross-linking pattern, however, remained unchanged, and in the HR infected cells the pattern of lower molecular weight multimers (≤hexamers) of E1 and E2 was uneffected when the glycoproteins were synthesized at 40°C. These results suggest that incubation of the cells at 40°C has not changed their permeability to the cross-linker, and that E1 and PE2 made at 40°C were more easily cross-linked due to increased aggregation. In the absence of any cross-linking, ts23 infected monolayers still had large amounts of SDS-stable E1 and PE2 aggregates (Fig. 10, panel G). Since monomeric E1 and PE2 were separated in the first dimension, heterodimers should be clearly distinguishable from homodimers. A comparison of E1 and PE2 aggregates in the dimer region from both cross-linked and noncross-linked monolayers, showed that the spots did not comigrate in the first dimension as would be expected for a pure population of heterodimers. In contrast to the pattern found for E1 and E2, this cross-linking pattern suggests preferential cross-linking of E1 and PE2 to themselves or possibly to unlabeled host cell proteins. Immunoprecipitation of noncross-linked monolayers infected with HR or ts23 with rabbit antisera against E1 or E2 (Fig. 11) showed that some of this aggregated PE2 was associated with E1. We also compared the amount of this PE2-E1 association in Sindbis HR infected cells grown at 37°C or 40°C with or without cross-linking (Fig. 12). Whether or not the monolayers were cross-linked, more PE2 could be precipitated with antiserum to E1 when the proteins were synthesized at 40°C than at 37°C (Fig. 12, lanes 2, 3, 5 and 6). The same result was found for monolayers solubilized in Triton X-100 (Fig. 12, lanes 7-12). Since these PE2-E1 aggregates could not be disrupted by Triton X-100 or SDS without reduction, some of them may be stabilized by disulfide bonds formed in vivo.

Fig. 11. Immunoprecipitation of noncross-linked HR and ts23 lysates. Noncross-linked lysates of HR infected (lanes 1-4) or ts23 infected (lanes 5-8) monolayers were prepared as described in Fig. 10 and precipitated with rabbit αE1 (lanes 2 and 6), αE2 (lanes 3 and 8), or nonimmune IgG (lanes 4 and 8) after preclearing with αC IgG. Samples of the lysates (lanes 1 and 5) and the immunoprecipitates were solubilized with SDS, reduced with 20 mM DTT and analyzed on a 10% slab gel (Laemmli, 1970).

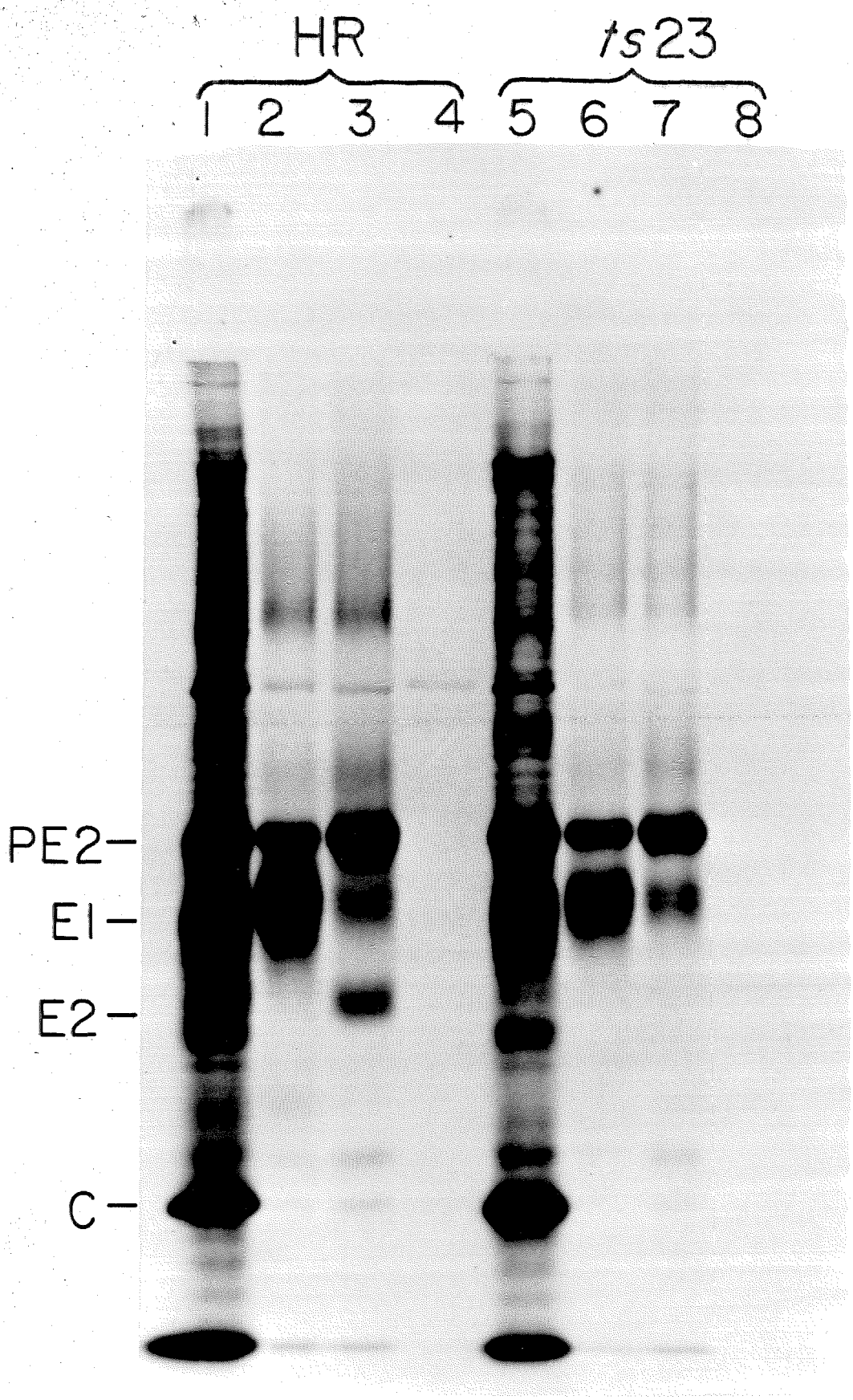
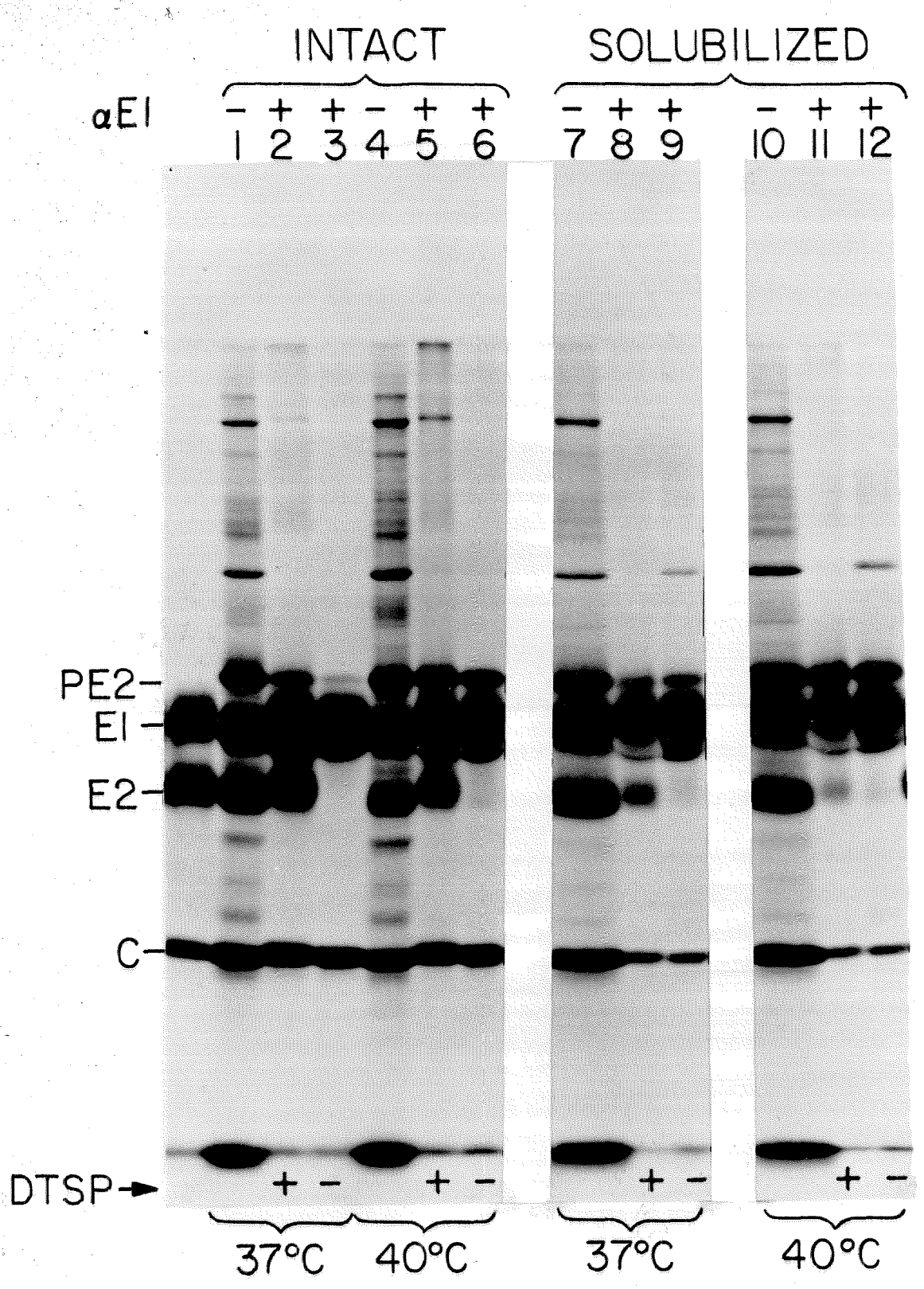

HR
ts23
1 2 3 4 5 6 7 8
PE2
E1
E2
C

Fig. 12. Association of PE2 and E1 synthesized at 37°C and 40°C. Confluent CEF monolayers were infected with Sindbis virus (HR strain) at 37°C, and labeled from 3-5 hr postinfection at either 37°C (lanes 1-3 and 7-9) or 40°C (lanes 4-6 and 10-12). Monolayers were washed with PBS and cross-linked with 0.2 mg/ml DTSP (lanes 2 and 5) or incubated in the presence of 2% DMSO as a control (lanes 3 and 6). Other monolayers were lysed in Triton X-100 (lanes 7-12) and cross-linked with 0.04 mg/ml DTSP (lanes 8 and 11), or incubated with 2% DMSO as a control (lanes 9 and 12). After quenching with lysine, samples were immunoprecipitated [indicated by (+)] with monoclonal αE1 IgG (lanes 2, 3, 5, 6, 8, 9, 11 and 12). Samples of the unprecipitated monolayers [indicated by (-)] are shown in lanes 1, 4, 7 and 10. Washed immunoprecipitates were resuspended in sample buffer containing 20 mM DTT, heated to 56°C for 30 min, and run on a 10% acrylamide slab gel (Laemmli, 1970). The far left lane is a purified virus marker.


INTACT
SOLUBILIZED
αE1
- + + - + +
1 2 3 4 5 6
- + +
7 8 9
- + +
10 11 12
PE2
E1
E2
C
DTSP
+ - + -
+ -
+ -
37°C
40°C
37°C
40°C

DISCUSSION

Bifunctional cross-linking reagents, whether cleavable or uncleavable, have been useful for the study of topological relationships between protein subunits in several virus systems (Garoff, 1974; Wiley et al., 1977; Dubovi and Wagner, 1977; Markwell and Fox, 1980). Our results with Sindbis virus show that in infected cells as well as mature virions, the E1 and E2 glycoproteins are closely associated (within a distance of 11 Å). Solubilization of the virion membrane with an excess of Triton X-100 produces glycoprotein-detergent 4.5 S complexes free of lipid (Helenius and Söderlund, 1973; Simons et al., 1973), which after stabilization by cross-linking could be shown to consist predominantly of E1-E2 heterodimers. The SFV membrane glycoproteins are also associated as heterodimers (Ziemiecki and Garoff, 1978), and in contrast to Sindbis virus, the interaction between SFV E1 and E2 cannot be disrupted by isoelectric focusing of the Triton X-100 solubilized glycoprotein complexes (Garoff et al., 1974; Dalrymple et al., 1976) In both viruses, the complex is largely disrupted by immunoprecipitation with rabbit antisera against the glycoproteins (Ziemiecki and Garoff, 1978; Bell et al., in preparation). In this regard, it is of interest that we have recently obtained a monoclonal antibody against Sindbis virus E1 which precipitates significant quantities of E2 and PE2 (J. Mayne and C. Rice, unpublished).

The interaction between E1 and E2 could play an important role during virus maturation. E1 contains at most a few amino acids on the cytoplasmic face of the host cell plasma membrane (Rice and Strauss, 1981; Rice et al., in preparation), and the cytoplasmic COOH-terminal portion of E2 (or PE2) is probably responsible for interacting with the nucleocapsid during budding. Thus, a stable E1-E2 complex in the plasma membrane would lead to the equimolar ratio of each structural protein found in mature virions (Garoff et al., 1974; Schlesinger et al., 1972), regardless of the ability of E1 to interact directly with the nucleocapsid.

When intact virions or infected cells were cross-linked we observed a lack of discrete glycoprotein aggregates larger than hexamers. These glycoprotein hexamers, which contain approximately equimolar amounts of E1 and E2, could represent higher-order viral subunits consisting of three E1-E2 heterodimers. Election microscopic examination of intact Sindbis virions has revealed a pattern of highly ordered hexagonal glycoprotein arrays which persists in viral membranes in which the nucleocapsid has been removed by treatment of the virions with low concentrations of Triton X-100 (von Bonsdorf and Harrison, 1978). If such an interaction between glycoprotein heterodimers occurs in the host cell plasma membrane leading to dense-packed higher-order glycoprotein aggregates, this "patching" would promote the exclusion of host cell glycoproteins (Strauss, 1978) as well as stable nucleocapsid binding.

Cross-linking of pulse-labeled monolayers (either intact or solubilized with Triton X-100) followed by immunoprecipitation with E1 antiserum demonstrated that some PE2 and E1 were associated as heterodimers. However, the finding that the majority of cross-linked, newly synthesized PE2 and E1 were not associated, suggested that these proteins were more readily cross-linked to themselves or other unlabeled host polypeptides (see below). Since only low levels of newly synthesized PE2 and E1 were cross-linked, such aggregates could represent aberrant forms of the glycoproteins not destined for incorporation into virions. The accumulation of high molecular weight PE2-E1 aggregates with longer chase periods lends support to this idea. It is therefore not possible to interpret these results as being clear-cut evidence for or against the existence of functional PE2-E1 complexes in infected cells. However, the different cross-linking behavior of PE2 and E2 suggests that they may have different orientations with respect to E1. In the case of SFV infected cells, newly synthesized E1 and PE2 can be solubilized into dimeric complexes with Triton X-100 which are capable of being cross-linked (Ziemiecki et al., 1980). While it has not been demonstrated directly that these dimers are PE2-E1 heterodimers, it seems likely (by analogy to the E1-E2

interaction) that functional PE2-E1 complexes may exist in infected cells. However, in the case of Sindbis virus the locations of the imidate reactive groups in PE2 and E1 do not favor their cross-linking with the reagents used in this study.

PE2 and E1 were extensively cross-linked into large aggregates when they were synthesized at 40°C instead of 37°C. This observation was true for the heat-resistant variant of Sindbis virus (HR) as well as two ts mutants defective in the PE2 cleavage, ts20 (complementation group E) and ts23 (complementation group D) (Strauss and Strauss, 1980). Lower concentrations of the cross-linking reagent produced discrete oligomers of E1 and PE2 which were primarily aggregates of these proteins cross-linked to themselves rather than to each other (although cross-linking to unlabeled host polypeptides could not be rigorously excluded). Even in the absence of chemical cross-linking, SDS-stable disulfide-linked aggregates of PE1 and E1 were found in HR and ts23 infected cells. Immunoprecipitation of these aggregates showed that more PE2 was associated with E1 at 40°C than at 37°C. Since only minor amounts of the aggregates were observed in pulse-chase experiments at 37°C, it is tempting to speculate that they represent aberrant glycoprotein precursors formed as a consequence of their synthesis of 40°C. Relevant to this discussion are several studies on the properties of viral glycoproteins synthesized in the presence of tunicamycin, an inhibitor of glycosylation. In the presence of tunicamycin, the multiplication of Sindbis virus and vesicular stomatitis virus is inhibited (Leavitt et al., 1977a). The nonglycosylated glycoproteins synthesized under these conditions do not migrate to the cell surface (PE2 is not cleaved to E2) and are found in large insoluble aggregates after solubilization with nonionic detergents such as Triton X-100 (Leavitt et al., 1977b). A study involving three strains of vesicular stomatitis virus and a ts mutant containing a lesion in the G glycoprotein demonstrated that differences in the sensitivity of these variants to tunicamycin was temperature-dependent and positively correlated with the intracellular aggregation of the G glycoprotein. Synthesis of the nonglycosylated G protein

at elevated temperatures led to an increased tendency towards aggregation (Gibson et al., 1979).

In Sindbis virus infected cells, our cross-linking results indicate that the glycoprotein precursors have this same tendency to aggregate at elevated temperatures even in the absence of tunicamycin. Mutations in E1 or PE2 increasing their tendency to aggregate at elevated temperatures and leading to the inhibition of virus multiplication may be the molecular basis for some of the ts lesions in complementation groups D and E (corresponding to the genes for E1 and PE2, respectively; for a review see Strauss and Strauss, 1980). At the restrictive temperature, the glycoproteins of mutants in complementation group D (ts10 and ts23) cannot be detected at the cell surface (Bell and Waite, 1977; Smith and Brown, 1977), and by immunofluorescence appear to be associated with rough endoplasmic reticulum (Saraste et al., 1980a). The glycoproteins of the ts1 mutant of SFV have a similar immunofluorescence pattern (Saraste et al., 1980b) and contain exclusively high-mannose type oligosaccharide chains at the restrictive temperature (Pesonen et al., 1981) which supports their localization in the rough endoplasmic reticulum. Upon shift to the permissive temperature (28-30°C) some of these glycoproteins can be transported to the plasma membrane (Saraste et al., 1980a,b) but are inefficiently incorporated into mature virions (Smith and Brown, 1977; Jones et al., 1977; Saraste et al., 1980b). Alternatively, the glycoproteins of ts20 (the only representative of group E) show some tendency to accumulate intracellularly, but are distributed in the rough endoplasmic reticulum, golgi, and the plasma membrane at the restrictive temperature (Bell and Waite, 1977; Smith and Brown, 1977; Saraste et al., 1980a). The cross-linking patterns of both ts20 and ts23 infected cells at the nonpermissive temperature indicate that PE2 and E1 are present as large aggregates, some of which may contain intermolecular disulfide bonds formed in vivo. The formation and stability of such aggregates at the restrictive temperature could affect the intracellular transport and glycosylation of the glycoprotein precursors as well as their

ability to be incorporated into mature virions after a shift to the permissive temperature. In future experiments, it will be of interest to determine whether a temperature-induced aggregation of the glycoproteins causes the temperature-sensitivity of these mutants or occurs merely as a byproduct of some other lesion affecting glycoprotein maturation.

ACKNOWLEDGEMENTS

We would like to thank Edith Lenches for preparing the CEF and BHK-21 monolayers. This work was supported by grant PCM 80-22830 from the National Science Foundation, and by grants GM 06965 and AI 10783 from NIH. C.M.R. was supported by Training Grant GM 00086 from NIH.

REFERENCES

Bell, J. R., Strauss, E. G. and Strauss, J. H. (1979) Purification and amino acid compositions of the structural proteins of Sindbis virus. Virology **97**, 287-294.

Bell, J. W. and Waite, M. R. F. (1977) Envelope antigens of Sindbis virus in cells infected with temperature-sensitive mutants. J. Virol. **21**, 788-791.

Birdwell, C. R. and Strauss, J. H. (1974) Replication of Sindbis virus IV. Electron microscopic study of the insertion of viral glycoproteins into the surface of infected chick cells. J. Virol. **14**, 366-374.

Bonatti, S., Cancedda, R. and Blobel, G. (1979) Membrane biogenesis: in vitro cleavage core glycosylation, and integration into microsomal membranes of Sindbis virus glycoproteins. J. Cell Biol. **80**, 219-224.

Bonner, W. M. and Laskey, R. A. (1974) A film detection method for tritium-labeled proteins and nucleic acids in polyacrylamide gels. Eur. J. Biochem. **46**, 83-88.

Bracha, M. and Schlesinger, M. J. (1976) Defects in RNA^+ temperature-sensitive mutants of Sindbis virus and evidence for a complex of PE2-E1 viral glycoproteins. Virology **74**, 441-449.

Brown, D. T. and Smith, J. F. (1975) Morphology of BHK-21 cells infected with Sindbis virus temperature-sensitive mutants in complementation groups D and E. J. Virol. **15**, 1262-1266.

Burge, B. W. and Pfefferkorn, E. R. (1966) Isolation and characterization of conditional-lethal mutants of Sindbis virus. Virology **30**, 204-213.

Cancedda, R., Villa-Komaroff, L., Lodish, H. F. and Schlesinger, M. (1975) Initiation sites for translation of Sindbis virus 42S and 26S messenger RNAs. Cell **6**, 215-222.

Dalrymple, J. M., Schlesinger, S. and Russell, P. K. (1976) Antigenic characterization of two Sindbis envelope glycoproteins separated by isoelectric focusing. Virology **69**, 93-103.

Davies, G. E. and Stark, G. R. (1970) Use of dimethyl suberimidate, a cross-linking reagent, in studying the subunit structure of oligomeric proteins. Proc. Natl. Acad. Sci. USA **66**, 651-656.

Dubovi, E. J. and Wagner, R. R. (1977) Spatial relationships of the proteins of vesicular stomatitis virus: induction of reversible oligomers by cleavable protein cross-linkers and oxidation. J. Virol. **22**, 500-509.

Dulbecco, R. and Vogt, M. (1954) One-step growth curve of western encephalomyelitis virus grown in vitro and analysis of the virus yields from single cells. J. Exp. Med. **99**, 183-199.

Garoff, H. (1974) Cross-linking of the spike glycoproteins in Semliki Forest virus with dimethylsuberimidate. Virology **62**, 385-392.

Garoff, H., Simons, K. and Renkonen, O. (1974) Isolation and characterization of the membrane proteins of Semliki Forest virus. Virology **61**, 493-504.

Garoff, H. and Söderlund, H. (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.

Garoff, H., Simons, K. and Dobberstein, B. (1978) Assembly of the Semliki Forest virus membrane glycoproteins in the membrane of the endoplasmic reticulum in vitro. J. Mol. Biol. **124**, 587-600.

Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H. and Delius, H. (1980a) Nucleotide sequence of the cDNA coding for the Semliki Forest virus membrane glycoproteins. Nature (London) **288**, 236-241.

Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H. and Delius, H. (1980b) The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino terminal region. Proc. Natl. Acad. Sci. USA **77**, 6376-6380.

Garvey, J. S., Cremer, N. E. and Sussdorf, D. H. (1977) "Methods in Immunology", W. A. Benjamin, Inc., Reading, Massachusetts, pp. 218-219.

Gibson, R., Schlesinger, S. and Kornfeld, S. (1979) The nonglycosylated glycoprotein of vesicular stomatitis virus is temperature-sensitive and undergoes intracellular aggregation at elevated temperatures. J. Biol. Chem. **254**, 3600-3607.

Helenius, A. and Söderlund, H. (1973) Stepwise dissociation of the Semliki Forest virus membrane with Triton X-100. Biochem. Biophys. Acta **307**, 287-300.

Jones, K. J., Scupham, R. K., Pfeil, J. A., Wan, K., Sagik, B. P. and Bose, H. R. (1977) Interaction of Sindbis virus glycoproteins during morphogenesis. J. Virol. **21**, 778-787.

Kääriäinen, L., Simons, K. and von Bonsdorff, C.-H. (1969) Studies in subviral components of Semliki Forest virus. Ann. Med. Exp. Fenn. **47**, 235-248.

Kessler, S. W. (1975) Rapid isolation of antigens from cells with a staphylococcal protein A-antibody absorbent: parameters of the interaction of antibody-antigen complexes with protein A. J. Immunol. **115**, 1617-1624.

Laemmli, U. K. (1970) Cleavage of the structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**, 680-685.

Laskey, R. A., and Mills, A. D. (1975) Quantitative film detection of ^{3}H and ^{14}C in polyacrylamide gels by fluorography. Eur. J. Biochem. **56**, 336-341.

Leavitt, R., Schlesinger, S. and Kornfeld, S. (1977a) Tunicamycin inhibits glycosylation and multiplication of Sindbis and vesicular stomatitis viruses. J. Virol. **21**, 375-385.

Leavitt, R., Schlesinger, S. and Kornfeld, S. (1977b) Impaired intracellular migration and altered solubility of nonglycosylated glycoproteins of vesicular stomatitis virus and Sindbis virus. J. Biol. Chem. **252**, 9018-9023.

Lomant, A. J. and Fairbanks, G. (1976) Chemical probes of extended biological structures: synthesis and properties of the cleavable protein cross-linking reagent [^{35}S] dithiobis (succinimidyl proprionate). J. Mol. Biol. **104**, 243-261.

Markwell, M. A. K. and Fox, C. F. (1980) Protein-protein interactions within paramyxoviruses identified by the native disulfide bonding or reversible chemical cross-linking. J. Virol. **33**, 152-166.

Nowinski, R. L., Lostrom, M. E., Tam, M. R., Stone, M. R. and Burnette, W. N. (1979) The isolation of hybrid cell lines producing monoclonal antibodies against the p15(E) protein of ecotropic murine leukemia viruses. Virology **93**, 111-126.

Ou, J.-H., Strauss, E. G. and Strauss, J. H. (1981) Comparative studies of the 3'-terminal sequences of several alphavirus RNAs. Virology **109**, 281-289.

Palacios, R., Palmiter, R. D. and Schimke, R. T. (1972) Identification and isolation of ovalbumin-synthesizing polysomes. I. Specific binding of ^{125}I-anti-ovalbumin to polysomes. J. Biol. Chem. **247**, 2316-2321.

Pesonen, M., Saraste, J., Hashimoto, K. and Kääriäinen, L. (1981) Reversible defect in the glycosylation of the membrane proteins of Semliki Forest virus ts1 mutant. Virology **109**, 165-173.

Pierce, J. S., Strauss, E. G. and Strauss, J. H. (1974) Effect of ionic strength on the binding of Sindbis virus to chick cells. J. Virol. **13**, 1030-1036.

Rice, C. M. and Strauss, J. H. (1981) Nucleotide sequence of the 26S mRNA of Sindbis virus and deduced sequence of the encoded virus structural proteins. Proc. Natl. Acad. Sci. USA, in press.

Saraste, J., von Bonsdorff, C.-H., Hashimoto, K., Karänen, S., Kääriäinen, L. (1980a) Reversible transport defects of virus membrane glycoproteins in Sindbis virus mutant infected cells. Cell Biol. Int. Rep. **4**, 279-286.

Saraste, J. von Bonsdorff, C.-H., Hashimoto, K., Kääriäinen, L., Karänen, S. (1980b) Semilki Forest virus mutants with temperature-sensitive transport defect of envelope proteins. Virology **100**, 229-245.

Schlesinger, S. and Schlesinger, M. J. (1972) Formation of Sindbis virus proteins: identification of a precursor for one of the envelope proteins. J. Virol. **10**, 925-932.

Schlesinger, M. J., Schlesinger, S. and Burge, B. W. (1972) Identification of a second glycoprotein in Sindbis virus. Virology **47**, 539-541.

Schlesinger, M. J. and Kääriäinen, L. (1980) "Translation and Processing of Alphavirus Proteins" in The Togaviruses, R. W. Schlesinger, Academic Press, New York, pp. 371-392.

Sefton, B. M. (1977) Immediate glycosylation of Sindbis virus membrane proteins. Cell **10**, 659-668.

Simmons, D. T. and Strauss, J. H. (1974) Translation of Sindbis virus 26S RNA and 49S RNA in lysates of rabbit reticulocytes. J. Mol. Biol. **86**, 397-409.

Simons, K., Helenius, A., and Garoff, H. (1973) Solubilization of the membrane proteins from Semliki Forest virus with Triton X-100. J. Mol. Biol. **80**, 119-133.

Smith, J. F. and Brown, D. T. (1977) Envelopment of Sindbis virus: synthesis and organization of proteins in cells infected with wild-type and maturation-defective mutants. J. Virol. **22**, 662-678.

Strauss, E. G. (1978) Mutants of Sindbis virus. III. Host polypeptides present in purified HR and ts103 virus particles. J. Virol. **28**, 466-474.

Strauss, E. G. and Strauss, J. H. (1980) "Mutants of alphaviruses: genetics and physiology" in The Togaviruses, R. W. Schlesinger, Academic Press, New York, pp. 393-426.

Strauss, J. H., and Strauss, E. G. (1977) "Togaviruses" in The Molecular Biology of Animal Viruses, D. Nayak, Dekker, New York, Vol. 1, pp. 111-166.

von Bonsdorff, C.-H. and Harrison, S. C. (1978) Hexagonal glycoprotein arrays from Sindbis virus membranes. J. Virol. **28**, 578-583.

Wiley, D. C., Skehel, J. J. and Waterfield, M. (1977) Evidence from studies with a cross-linking reagent that the hemagglutinin of influenza virus is a trimer. Virology **79**, 446-448.

Wirth, D. F., Katz, F., Small, B. and Lodish, H. F. (1977) How a single Sindbis virus mRNA directs the synthesis of one soluble and two integral membrane glycoproteins. Cell **10**, 253-263.

Yin, F. H. (1969) Temperature-sensitive behavior of the hemagglutinin in a temperature-sensitive mutant virion of Sindbis. J. Virol. **4**, 547-548.

Ziemiecki, A. and Garoff, H. (1978) Subunit composition of the membrane glycoprotein complex of Semliki Forest virus. J. Mol. Biol. **122**, 259-269.

Ziemiecki, A., Garoff, H. and Simons, K. (1980) Formation of the Semliki Forest virus membrane glycoprotein complexes in the infected cell. J. gen. Virol. **50**, 111-123.