THE ISOLATION AND PARTIAL BIOCHEMICAL ANALYSIS OF SINDBIS VIRUS PROTEINS

Thesis by

John Richard Bell

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1981

(Submitted January 6, 1981).

Abstract

Recently, the amino acid sequences have been reported for several proteins, including the envelope glycoproteins of Sindbis virus, which all probably span the plasma membrane with a common topology: a large N-terminal, extracellular portion, a short region buried in the bilayer, and a short C-terminal intracellular segment. The regions of these proteins buried in the bilayer correspond to portions of the protein sequences which contain a stretch of hydrophobic amino acids and which have other common characteristics, as discussed. Reasons are also described for uncertainty, in some proteins more than others, as to the precise location of some parts of the sequence relative to the membrane.

The signal hypothesis for the transmembrane translocation of proteins is briefly described and its general applicability is reviewed. There are many proteins whose translocation is accurately described by this hypothesis, but some proteins are translocated in a different manner.

The transmembraneous glycoproteins E1 and E2 of Sindbis virus, as well as the only other virion protein, the capsid protein, were purified in amounts sufficient for biochemical analysis using sensitive techniques. The amino acid compositon of each protein was determined, and extensive N-terminal sequences were obtained for E1 and E2. By these techniques E1 and E2 are indistinguishable from most water soluble proteins, as they do not contain an obvious excess of hydrophobic amino acids in their N-terminal regions or in the intact molecule.

The capsid protein was found to be blocked, and so its N-terminus could not be sequenced by the usual methods. However, with the use of a special labeling technique, it was possible to incorporate tritiated acetate into the N-terminus of the protein with good specificity, which was useful in the purification of peptides from which the first amino acids in the N-terminal sequence could be identified.

Nanomole amounts of PE2, the intracellular precursor of E2, were purified by an immuno-affinity technique, and its N-terminus was analyzed. Together with other work, these results showed that PE2 is not synthesized with an N-terminal extension, and the signal sequence for translocation is probably the N-terminal amino acid sequence of the protein. This N-terminus was found to be 80-90% blocked, also by N-acetylation, and this acetylation did not affect its function as a signal sequence. The putative signal sequence was also found to contain a glycosylated asparagine residue, but the inhibition of this glycosylation did not lead to the cleavage of the sequence.

Table of Contents

Page

Chapter 1 The Structure and Synthesis of the Envelope Glycoproteins of Sindbis Virus and their Relationship to Other Membrane Proteins 1

Chapter 2 Purification and Amino Acid Compositions of the Structural Proteins of Sindbis Virus 20

Chapter 3 Amino-terminal Sequence Analysis of the Structural Proteins of Sindbis Virus 29

Chapter 4 In Vivo N-terminal Acetylation of Sindbis Virus Proteins 35

Chapter 5 The N-Terminus of PE2 in Sindbis Virus Infected Cells 63

Chapter 1

The Structure and Synthesis of the Envelope Glycoproteins of Sindbis Virus and Their Relationship to Other Membrane Proteins

The lipid-containing membranes of a cell provide an effective barrier to the movement of proteins and small water soluble molecules, preventing the loss of the cellular contents to the environment and providing for the compartmentalization of a variety of activities within the cell. However, the cell requires proteins which, while maintaining their association with the cell, must be largely external and so able to interact with the extracellular environment, functioning as receptors for extra-cellular chemical signals or mediating cell-cell interactions. This implies that a mechanism must exist whereby the extracellular portions of these proteins are selectively translocated across the membrane during or after their synthesis. Proteins secreted by the cell, or proteins synthesized in the cytoplasm but required inside other cellular compartments such as mitochondria, must also cross membrane barriers.

The membrane proteins of certain enveloped viruses, such as Sindbis virus (and another alphavirus, the closely related Semliki Forest virus), vesicular stomatitus virus, and influenza virus, have been extensively studied and appear to be very similar in many respects to integral membrane proteins of the host cell plasma membrane (see below). These viruses acquire their lipid envelope by budding through a region of the host cell plasmalemma containing the viral membrane proteins (Compans and Klenk, 1979), in such a way that the topological arrangement of these proteins is much the same in the virion as in the plama membrane. Thus, the interior of the lipid bilayer of the virus corresponds to the cytoplasmic side of the plasma membrane. Since host membrane proteins are excluded from the membrane during budding, the study of the arrangement of the viral envelope protein is not complicated by the presence of a large number of different proteins in the virion. In the following two sections, which discuss the structure and translocation of certain integral membrane proteins, the usefulness of viral systems for the study of these proteins is evident.

Structure

In Sindbis virus, the topological arrangement of the two envelope glycoproteins, E1 and E2, has been studied by treating the virus particle with protease (Compans, 1971; Garoff and Soderlund, 1978). Most of the mass of these proteins is digested, but the regions of the proteins which are either buried in, or internal to, the membrane are protected, and these regions (called roots) have been isolated for biochemical characterization, including protein sequenation (C. M. Rice and M. W. Hunkapiller, personal communication). In addition, the complete amino acid sequences of E1 and E2 have been deduced from the nucleotide sequence of their mRNA (Rice and Strauss, manuscript in preparation). The results of these studies show that the N-termini of both E1 and E2 are external to the lipid bilayer of the viral envelope, and both are anchored in the bilayer by a hydrophobic sequence, a linear sequence of 23-33 hydrophobic residues, at or near the C-terminal end of the protein. In the case of E2, the protein must completely span the bilayer, since approximately 30 amino acids can be removed from the C-terminal end by proteolytic attack at the cytoplasmic side of the membrane in microsomal vesicles (Wirth et al., 1977). The location of the C-terminal end of this 30-amino acid segment is unknown, but its primary structure (unpublished observations; Rice and Strauss, manuscript in preparation) suggests the interesting possibility that it may loop back into the bilayer, although there is no direct evidence for this arrangement. The C-terminus of E1 probably also spans the membrane, but with only 2-5 residues exposed on the cytoplasmic side.

Semliki Forest virus, closely related to Sindbis virus, has been studied in a similar manner (Garoff et al., 1978, 1980; Garoff and Soderlund, 1978), although the characterization of the roots to a similar extent has not been described. The topological orientations of the Semliki Forest virus membrane glycoproteins E1 and E2 are identical to the results described above for Sindbis virus E1 and E2, respectively.

The key features of the approach taken to determine the orientation of the Sindbis virus E1 and E2 glycoproteins in the membrane are the use of a probe (protease in these experiments) for the external, N-terminal, portion of the molecule, use of protease as a probe for the C-terminal region of E2 internal to the bilayer, and the isolation and characterization of the roots. Since only two roots are present after protease digestion of the virus particle, they can be recovered for analysis relatively easily. A similarly complete analysis has been performed for only one other membrane proteins, the G protein of vesicular stomatitus virus (Rose et al., 1980; Katz et al., 1977) and its orientation is the same as that of Sindbis virus E2. Most of the protein, including the N-terminus, is external to the membrane of the virus particle; there is a short hydrophobic segment near the C-terminal end which anchors the protein in the membrane, and there is an additional short cytoplasmic segment at the extreme C-terminal end.

Several other membrane proteins, including glycophorin A (Segrest et al., 1973; Bretscher et al., 1971) and the HLA-A and HLA-B human histocompatibility antigens (Walsh and Crumpton, 1977; Pober et al., 1978) have been directly shown to traverse the membrane. In the case of glycophorin A, a hydrophobic sequence has been found at the expected position (near the C-terminal end) in the amino acid sequence of the complete molecule (Tomita and Marchesi, 1975; Tomita et al., 1978). A C-terminal portion of the HLA protein has been sequenced, which includes the C-terminal two residues of a hydrophobic sequence and extends to the end of the molecule (Robb et al., 1978).

An association with the membrane, although less precisely defined experimentally, has been shown for a number of other membrane proteins, including the HA protein of influenza virus (Skehel and Waterfield, 1975) (as well as a variety of other viral membrane proteins [see Compans and Klenk, 1979]). This association is demonstrated by the observation that it is possible to solubilize a large portion of the

molecule by removing a small C-terminal fragment by limited proteolytic digestion, whereas detergent is required to solubilize the intact protein. The presence of an intramembraneous region has been shown by the use of a lipid soluble labeling reagent (Goldman et al., 1979). Limited proteolytic cleavage of μ_M, the heavy chain of membrane-bound IgM, gives similar results (Melcher et al., 1975). In this case there are two forms of the protein, and a correlation is found between the C-terminal sequence of the protein and whether the protein is soluble or membrane-bound (Kehry et al., 1980; Rogers et al., 1980). The HA of two strains of influenza (Jou et al., 1980; Porter et al., 1979) and the μ_M chain (Rogers et al., 1980) have been sequenced (as cDNA), and both contain hydrophobic segments near their C-terminal ends.

Some features of the hydrophobic sequences of these nine proteins, which are known or presumed to contain a region buried in the membrane bilayer, are presented in Table 1. The sequences all contain certain distinctive features. All are relatively short, ranging in length from 20 to 33 residues, and all consist exclusively of hydrophobic residues (including Ser and Thr). Certain amino acids, Asp, Glu, Asn, Gln, His, Lys, Arg, and Pro, are never found in the interior of these sequences, and these amino acids serve to delineate the ends of the hydrophobic sequences. The transmembrane hydrophobic sequences always end at their C-terminal ends with a basic group, and there is almost always one or more additional basic residues among the next few amino acids. However, there are fewer constraints on the residues which terminate the sequences at the other end. There is almost always one or more charged groups, often consisting of clusters of acidic amino acids, at or near the N-terminal end of the hydrophobic sequence, but in one case there appears to be no charged group in this region of the molecule at all (Semliki Forest virus E2, but see below).

One other interesting feature of these hydrophobic sequences is that they are all quite close to the end of the protein, for there are only 2 to 33 additional residues C-terminal to these regions. In addition to whatever this observation may suggest

Table 1

Structure of Hydrophobic Sequences

Protein[a]	Number of residues in hydrophobic sequence	Number of residues C-terminal to hydrophobic sequence	Nearby charged residues[b]	
			N-terminal	C-terminal
SV E1	33	2	K(1)	R(1-2)
SV E2	26	33	H(1),R(2),H(3)	K(1),R(3-4),E(5)
SFV E1	24	2	K(1)	R(1-2)
SFV E2	28	31	None	R(1-2),K(3)
VSV G	20	29	K(1)	R(1),H(5)
Flu HA	26	11	D(1),K(2)	K(1 or 2)[c]
Glycophorin A	23	36	E(1),E(3)	R(1-2),K(5)
HLA-A, HLA-B	?	30	?	R(1-2),K(3)
μ_M	25	3	E(2),E(5)	K(1),K(3)

[a]Abbreviations: SV = Sindbis Virus, SFV = Semliki Forest Virus, VSV = vesicular stomatitus virus. See text for references.

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

[c]Two strains.

about the function or synthesis of these regions of the proteins, this can present practical problems in the demonstration of the transmembraneous nature of these proteins. In the case of the influenza HA protein, Sindbis and Semliki Forest virus E1, and μ_M, the assignment of the C-terminus of the molecule to the cytoplasmic side of the membrane is based largely upon the mechanism of the synthesis and insertion of the protein into the membrane (see below).

It is presumed that the intramembranous regions of these nine proteins, the regions actually buried in the interior of the lipid bilayer, consist of the hydrophobic sequences described above. This is based on the extremely hydrophobic character of the sequences and on their location in the linear sequence in the molecule (i.e., close to the C-terminal end) to which the intramembranous region has been assigned by other means. However, this localization of the intramembranous region in the linear sequence of the protein by chemical or enzymatic probes is not as precise as one would like. The most exact localization is obtained by examining the sequences of the protease resistant roots, but even in these cases the roots begin 5 (Sindbis virus E2) (Rice and Strauss, manuscript in preparation) to 14 (vesicular stomatitus virus G protein) (Rose et al., 1980) residues from the N-terminal end of the hydrophobic sequence. C-terminal localization is even more imprecise, in the case of those proteins which extend for some distance into the cytoplasm, for it is based on the difference of the estimated molecular weights of the intact protein and of the protein with its cytoplasmic C-terminus removed.

In most cases the identification of the ends of the intramembranous regions in the protein sequence can be based with reasonable confidence on the presence of charged amino acids (which are not expected to be in a non-polar environment) at the ends of the hydrophobic sequences and the requirement for at least a minimum number of hydrophobic amino acids to span the bilayer. However, a limitation of this approach is illustrated by the case of Semliki Forest virus E2, in which the proposed

identification of the N-terminal extent of the intramembranous region is based on the presence of a proline residue in the sequence and by a comparison with Sindbis virus E2. N-terminal to the proline are five hydrophobic amino acids, then a gluatmine residue, and then three more hydrophobic residues, whose locations relative to the bilayer are unknown.

An additional interesting observation can be made from a comparison of the sequences of the Sindbis (Rice and Strauss, manuscript in preparation) and Semliki Forest virus (Garoff et al., 1980) E1 glycoproteins. These two viruses are quite similar to the extent that the overall sequence homology of the structural proteins and their precursors is 47%, another 12% of the residues are conservative substitutions, and the sequences can be easily aligned (Rice and Strauss, manuscript in preparation). Nonetheless, the hydrophobic sequence of the Sindbis protein is eight residues longer than that of the Semliki Forest virus protein, with this extension at the N-terminal end of the aligned sequences. There would seem to be only two possible explanations for this observation. The first is that such similar viruses have rather different conformations in this region of the protein, both in the intramembranous segment and in the adjacent portion of the extramembranous region. The alternative explanation is that the intramembranous segments of the proteins are not located in the amino acid sequences precisely as predicted.

Translocation

The first observations on a molecular level of the translocation of a protein across a membrane were obtained during studies on a secreted protein (IgG) in eukaryotes (Milstein et al., 1972), and these and other studies (Blobel and Dobberstein, 1975a) formed the basis for the signal hypothesis advanced by Blobel and Dobberstein (1975b). Much additional work has led to some elaboration of this hypothesis (Blobel et al., 1979) and its confirmation in many cases, but it clearly does not accurately describe the processing of all proteins which are translocated across

membranes (Blobel et al., 1979; Davis and Tai, 1980; Wickner, 1980), as discussed below.

The important features of the signal hypothesis (Blobel et al., 1979) may be summarized briefly as follows. Secreted proteins are synthesized on the membrane bound ribosomes of the rough endoplasmic reticulum (Palade, 1975), and the initial interaction with the membrane is mediated by a signal sequence, a 15-30 amino acid extension at the N-terminal end of the protein. This interaction, which can occur for only a short time after the synthesis of the signal sequence, while most of the rest of the protein is untranslated, results in an additional, salt-sensitive binding of the ribosome to protein components of the membrane, the formation of a protein pore (or the opening of a gate) through the bilayer, and the cotranslational insertion of the protein through the pore into the lumen of the endoplasmic reticulum. Shortly after its translocation the signal sequence is cleaved off from the rest of the protein, and the pore disassembles (or the gate closes) when translocation is completed.

Some of the protein components of the membrane which are involved in this activity have been identified and partially characterized. The signal peptidase responsible for cleaving the signal sequence has been partially purified and its activity has been examined (Jackson and Blobel, 1977), an active proteolytic fragment of a protein necessary for translocation has been identified (Walter et al., 1979), and two membrane proteins of the endoplasmic reticulum, called ribophorins, which bind ribosomes have been found (Kreibich et al., 1978a,b). However, as yet no protein has been found which can be shown to form the pore, or gate, in the membrane. The portion of the system which resides on the protein to be translocated, in contrast, has been extensively studied, and the complete or partial amino acid sequences of the signal sequences of a number of secreted proteins have been determined (see Blobel, 1979). Although these signal sequences share certain similarities, the lack of direct sequence homology makes it difficult to draw conclusions about the mechanisms

involved solely from amino acid sequence information. Apparently secondary and tertiary structure is important in the function of a signal sequence (Blobel, 1979).

All of the signal sequences do contain a stretch of at least 9 residues which are hydrophobic, leading to speculation concerning a possible interaction of these sequences with the hydrophobic interior of the membrane. Cleavage by signal peptidase occurs at peptide bonds C-terminal to glycine, alanine, serine, threonine, or cysteine, amino acids with small side chains, but of course there are many other such bonds in signal sequences which are not cleaved, and it is not known what other structural features distinguish the cleavage site. Proline or glycine, or both, are often found about five residues N-terminal to the cleavage site, suggesting that β-bends, which often contain these residues (Chou and Fasman, 1974), are important in the formation of the cleavage site or in the conformation of the signal sequence during translocation (Inouye and Halegoua, 1980). However, it is difficult to predict such structures from the amino acid sequence of a protein, and there are signal sequences which do not contain these two residues.

The model for the translocation of secretory proteins described above has been extended to explain the translocation of portions of integral membrane proteins (Blobel and Dobberstein, 1975b), the only modification being the postulated existence of a stop transfer sequence in the membrane protein. This sequence would cause the dissolution of the pore in the membrane, stopping translocation at this point and leaving the rest of the polypeptide chain embedded in the lipid bilayer, and possibly extending some distance into the cytoplasm. Several well studied plasma membrane proteins, the G protein of vesicular stomatitus virus (Lingappa et al., 1979), the HA protein of influenza virus (Elder et al., 1979), and the histocompatibility antigens H-2D of mouse (Dobberstein et al., 1979) and HLA-A and HLA-B of man (Ploegh et al., 1979), have been shown to conform to this model. The initial events in the translocation of these proteins are identical to the events described above for

secretory proteins and the signal sequences of the two viral proteins are indistinguishable from those found in secreted proteins (Rose et al., 1980; Jou et al., 1980). The translocation of E2 of Sindbis virus is also similar, but the signal sequence is not cleaved and it contains some unusual structural features, as discussed in detail in Chapter 5 of this thesis.

Membrane proteins whose polypeptide chains traverse the membrane several times are proposed to contain alternating stop transfer and internal signal sequences (Blobel et al., 1979). Precedent for the latter comes from studies on an unusual secreted protein, ovalbumin. This protein is not translated with an N-terminal extension (Palmiter et al., 1978), and the signal sequence is found in the C-terminal portion of the molecule and can be identified since it retains its activity to some extent in the intact molecule upon denaturation or when the peptide is isolated as a tryptic fragment (Lingappa et al., 1979).

The transmembrane translocation of proteins destined for the interior of various membrane-bound organelles has been studied, and in some cases the scheme utilized by secretory proteins is also followed by these proteins. Thus, lysozomal enzymes (Erickson and Blobel, 1979), and probably also intergral membrane proteins of the peroxizome (Goldman and Blobel, 1978), are translocated across the membrane of the rough endoplasmic reticulum as they are synthesized. In contrast with this co-translational translocation is the post-translational translocation observed for certain proteins made in the cytoplasm but used in the mitochondria or chloroplast. The small subunit of ribulose-1,5-biphosphate carboxylase is synthesized with an N-terminal extension but on free, not membrane-bound, ribosomes (Dobberstein et al., 1977). The post-translational translocation of the precursor, and the cleavage of its N-terminal extension, can be observed upon incubation with isolated chloroplasts in vitro (Highfield and Ellis, 1978). A similar situation has been found in studies of the α, β, and γ subunits of yeast mitochondrial F_1-ATPase. Translocation is again post-

translational and is accompanied by the cleavage of an N-terminal extension of the mature form of the protein about 60 amino acids long (Maccecchini et al., 1979). Other workers showed that this translocation requires ATP (Nelson and Schatz, 1979).

The machinery for translocating proteins across membranes has evolved very little over time, as demonstrated by the correct translocation and cleavage of the signal sequence of proinsulin coded by fish mRNA and synthesized in a system consisting of wheat germ ribosomes and dog pancreatic membranes (Shields and Blobel, 1977). Perhaps even more interesting is the finding that the insertion of certain membrane proteins into the inner and outer membrane of Escherichia coli, and the translocation of some proteins into the periplasmic space, is functionally identical to the translocation across the membrane of the endoplasmic reticulum in mammalian cells, to the extent that the features of bacterial signal sequences cannot be distinguished from those of their eukaryotic counterparts. However, as in plants and higher animals, there are several examples of post-translational translocation in bacteria (for reviews, see Emr et al., 1980; Inouye and Halegoua, 1980; Wickner, 1980).

Conclusions

The elucidation of the complete amino acid sequences of a number of membrane proteins which traverse the membrane has been a major accomplishment. The distinctive common features which are found, some obviously related to function, are important for the confirmation of our understanding of the forces involved in the construction of the membrane, particularly with respect to its protein components. However, additional work is necessary to resolve unanswered questions, such as the conformation (and the allowed variability in the conformation) of the polypeptide chain in the lipid bilayer and the role of the charged residues at the ends of these sequences in the synthesis and integrity of these membrane proteins. It will also be interesting to compare these sequences with those of proteins such as band 3 of the

erythrocyte, which forms an anion transport channel in the membrane and which has a larger fraction of its mass in the interior of the bilayer (Rao, 1979; Fukuda et al., 1978).

Despite their sometimes major differences, there are two common features of the translocation of proteins across membranes. First, protein components of the membrane are required, at least in eukaryotes, as these have been isolated in the case of co-translational translocation and must be present to provide a sorting mechanism in the case of post-translational translocation. Second, the information necessary for this translocation is contained in the sequence of the protein, rather than in some structural feature of the mRNA or due to translation by a special class of ribosomes. This point is most directly demonstrated by the observation that in the case of Sindbis virus, a single mRNA with a single initiation site for translation encodes the three structural proteins of the virus (Cancedda et al., 1975). The primary translation product of this mRNA is processed by proteolytic cleavage to yield both the capsid protein, which is released into the cytosol, and two membrane proteins which are co-translationally inserted into the lumen of the rough endoplasmic reticulum (Wirth et al., 1977; Garoff et al., 1978). Nonetheless, the differences between the signal sequences which effect translocation across membranes are remarkable, particularly when compared to the similarities found between the intramembraneous regions of membrane proteins.

References

Blobel, G. and Dobberstein, B. (1975a). Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. J. Cell Biol. **67**, 835-851.

Blobel, G. and Dobberstein, B. (1975b). Transfer of proteins across membranes. II. Reconstitution of functional rough microsomes by heterologous components. J. Cell Biol. **67**, 852-862.

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.

Bretscher, M. S. (1971). Major human erythrocyte glycoprotein spans the cell membrane. Nature NewBiol. **231**, 229-232.

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

Chou,P. Y., and Fasman, G. d. (1974). Prediction of protein conformation. Biochemistry **13**, 222-245.

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

Compans, R. W., and Klenk, H.-D. (1979). Viral membranes. In "Comprehensive Virology" (H. Fraenkel-Conrat and R. R. Wagner, eds.), Vol. 13, pp. 293-407. Plenum Press, New York.

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

Dobberstein, B., Blobel, G. and Chua, N. H. (1977). In vitro synthesis and processing of a putative precursor for the small subunit of ribulose-1,5-bisphosphate carboxylase of Chlamydomonas reinhardtii. Proc. Nat. Acad. Sci. USA **74**, 1082-1085.

Dobberstein, B., Garoff, H., Warren, G., and Ribonson, P. J. (1979). Cell-free synthesis and membrane insertion of mouse H-2D histocompatibility antigen and β_2-microglobulin. Cell **17**, 759-769.

Elder, K. T., Bye, J. M., Skehel, J. J., Waterfield, M. D., and Smith, A. E. (1979). In vitro synthesis, glycosylation, and membrane insertion of influenza virus haemagglutinin. Virology **95**, 343-350.

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.

Erickson, A. H., and Blobel, G. (1979). Early events in the biosynthesis of the lysozomal enzyme cathepsin D. J. Biol. Chem. **254**, 11771-11774.

Fukuda, M., Eshdat, Y., Tarome, G., and Marchesi, V. T. (1978). Isolation and characterization of peptides derived from the cytoplasmic segment of Band 3, the predominant intrinsic membrane protein of the human erythrocyte. J. Biol. Chem. **253**, 2419-2428.

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

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., and Soderlund, 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.

Goldman, B. M., and Blobel, G. (1978). Biogenesis of peroxisomes: intracellular site of synthesis of catalase and uricase. Proc. Nat. Acad. Sci. USA **75**, 5066-5070.

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 **280**, 841-843.

Highfield, P. E. and Ellis, R. J. (1978). Synthesis and transport of the small subunit of chloroplast ribulose bisphosphate carboxylase. Nature, London **271**, 420-424.

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

Jackson, R. C., and Blobel, G. (1977). Post-translational cleavage of presecretory proteins with an extract of rough microsomes from dog pancreas containing signal peptidase activity. Proc. Nat. Acad. Sci. USA **74**, 5598-5602.

Jou, W. M., Verhoeyen, M., Devon, R., Saman, E., Fang, R., Huylebroeck, D., Fiers, W., Threlfall, G., Barber, C., Carey, N., and Entage, 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. Nat. Acad. Sci. USA **74**, 3278-3282.

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

Kreibich, G., Ulrich, B. C. and Sabatini, D. D. (1978a). Proteins of rough microsomal membranes related to ribosome binding. I. Identification of ribophorins I and II, membrane proteins characteristic of rough microsomes. J. Cell Biol. **77**, 464-487.

Kreibich, G., Freienstein, C. M., Pereyra, P. N., Ulrich, B. C. and Sabatini, D. D. (1978b). Proteins of rough microsomal membranes related to ribosome binding. II. Cross-linking of bound ribosomes to specific membrane proteins exposed at the binding sites. J. Cell Biol. **77**, 488-505.

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

Maccecchini, M. L., Rudin, Y., Blobel, G., and Schatz, G. (1979). Import of proteins into mitochondria. Precursor forms of the extramitochondrially-made F_1-ATPase subunits in yeast. Proc. Natl. Acad. Sci. USA **76**, 343-347.

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

Milstein, C., Brownlee, G. G., Harrison, T. M., and Mathews, M. B. (1972). A possible precursor of immunoglobulin light chains. Nature New Biol. **239**, 117-120.

Nelson, N., and Schatz, G. (1979). Energy-dependent processing of cytoplasmically made precursors to mitochondrial proteins. Proc. Nat. Acad. Sci. USA **76**, 4365-4369.

Palade, B. E. (1975). Intracellular aspects of the process of protein secretion. Science **189**, 347-358.

Palmiter, R. D., Gagnon, J., and Walsh, K. A. (1978). Ovalbumin: a secreted protein without a transient hydrophobic leader sequence. Proc. Nat. Acad. Sci. USA **75**, 94-98.

Ploegh, H. L., Cannon, L. E., and Strominger, J. L. (1979). Cell-free translation of the mRNAs for the heavy and light chain of HLA-A and HLA-B antigens. Proc. Nat. Acad. Sci. USA **76**, 2273-2277.

Pober, J. S., Guild, B. C., and Strominger, J. L. (1978). Phosphorylation in vivo and in vitro of human histocompatibility antigens (HLA-A and HLA-B) in the carboxy-terminal intracellular domain. Proc. Nat. Acad. Sci. USA **75**, 6002-6006.

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

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

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 chain. Cell **20**, 303-312.

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

Segrest, J. P., Kahane, I., and Marchesi, V. T. (1973). Major glycoprotein of the human erythrocyte membrane: evidence for an amphipathic molecular structure. Arch. Biochem. Biophys. **155**, 167-183.

Shields, D., and Blobel, G. (1977). Cell-free synthesis of fish preproinsulin, and processing by heterologous mammalian microsomal membranes. Proc. Nat. Acad. Sci. USA **74**, 2059-2063.

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

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.

Tomita, M., and Marchesi, V. T. (1975). Amino-acid sequence and oligosaccharide attachment sites of human erythrocyte glycophorin. Proc. Nat. Acad. Sci. USA **72**, 2964-2968.

Walsh, F. S., and Crumpton, M. J. (1977). Orientation of cell-surface antigens in the lipid bilayer of lymphocyte plasma membrane. Nature **269**, 307-311.

Walter, P., Jackson, R. C., Marcus, M. M., Lingappa, V. R., and Blobel, G. (1979). Tryptic dissection of receptor(s) for the translocation of nascent presecretory proteins across the microsomal membrane and reconstitution of translocation activity. Proc. Nat. Acad. Sci. USA **76**, 1795-1977.

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.

Chapter 2

Purification and Amino Acid Compositions of the Structural Proteins of Sindbis Virus

John R. Bell, Ellen G. Strauss and James H. Strauss

This chapter was published in Virology

VIROLOGY 97, 287–294 (1979)

Purification and Amino Acid Compositions of the Structural Proteins of Sindbis Virus

JOHN R. BELL, ELLEN G. STRAUSS, AND JAMES H. STRAUSS[1]

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

Accepted May 12, 1979

The envelope glycoproteins E1 and E2 of Sindbis virus (a member of the alphavirus group of the Togavirus family) have been purified on the basis of their differential binding to glass wool in the presence of the nondenaturing detergent Triton X-100, and milligram amounts of all three structural proteins of the virus have been prepared. The amino acid compositions of these proteins have been determined for two strains of the virus, *wt* and HR. Although it is clear from other studies that one or more structural proteins of the heat-resistant variant have been altered relative to the wild type, we were unable to detect such changes at the level of amino acid composition. As expected, the capsid protein is rich in the basic amino acids, and in this respect shows a strong similarity to the capsid protein of the closely related Semliki Forest virus. The molecular weight of the capsid protein has been estimated from the composition data to be 30,000 ± 600 daltons, in good agreement with the results obtained by sodium dodecyl sulfate–polyacrylamide gel electrophoresis. The envelope proteins E1 and E2 contain a proportion of polar amino acids which is typical of water-soluble proteins, with no apparent excess of amino acids with hydrophobic side chains. Thus it is unlikely that extensive regions of these proteins are buried in the hydrophobic interior of the lipid bilayer of the viral envelope.

INTRODUCTION

Sindbis virus is an enveloped virus with a particularly simple protein composition. Of the three structural proteins, one, the capsid protein, is found associated with the viral RNA in a nucleocapsid structure (Strauss *et al.*, 1968), while the other two are glycoproteins (Strauss *et al.*, 1970; Schlesinger *et al.*, 1972) and are largely external to the lipid of the envelope (Compans, 1971). The glycoproteins, E1 and E2, are anchored in the lipid bilayer by short hydrophobic segments (Garoff and Söderlund, 1978) and one or both extends completely through the membrane (Garoff and Simons, 1974). In the infected cell, nucleocapsids are assembled in the cytoplasm, while the envelope glycoproteins appear to be synthesized and glycosylated by the host cell's normal membrane protein synthesizing system (Wirth *et al.*, 1977) and are found at the plasma membrane (Birdwell and Strauss, 1974). The final maturation of the virus particle occurs as a preformed nucleocapsid buds through the plasmalemma, thus acquiring the envelope of lipid and E1 and E2 (Strauss and Strauss, 1977). Host cell membrane proteins are efficiently excluded from the budding virus and are not found in the virion (Pfefferkorn and Clifford, 1964; Strauss, 1978).

We are currently studying the structure of the Sindbis virus proteins and have recently reported the *N*-terminal amino acid sequences of E1 and E2 (Bell *et al.*, 1978). In this paper, a method is presented for the large scale purification of the envelope glycoproteins in the nondenaturing detergent Triton X-100 and the amino acid compositions of the three structural proteins are reported.

MATERIALS AND METHODS

Virus strains. Virus strains used were the wild type (from Dr. B. Burge) and HR (large plaque) strain (Burge and Pfefferkorn, 1966).

[1] To whom reprint requests should be addressed.

0042-6822/79/120287-08$02.00/0
Copyright © 1979 by Academic Press, Inc.
All rights of reproduction in any form reserved.

Growth and purification of unlabeled virus. Roller bottles (750 cm^2) of chick embryo fibroblasts were prepared as previously described (Pierce *et al.*, 1974) and infected at 37° at a multiplicity of about 0.01. The culture fluid (40 ml) was harvested 18 hr after infection and stored frozen at −70°. Virus was concentrated from pooled culture fluid by adding 0.25 vol of 40% polyethylene glycol (average molecular weight 6000 to 7500) in 2 *M* NaCl (Pierce *et al.*, 1974). After 1 hr at 0°, the precipitated virus was collected by centrifugation at 10,000 rpm for 45 min, in a Sorvall GSA rotor, the pelleted virus was resuspended in a small volume (about 25 ml/liter of culture fluid) of TNE (0.2 *M* NaCl, 0.05 *M* Tris, pH 7.4, 1 m*M* EDTA) and insoluble material was removed by centrifugation at 2000 rpm for 10 min in the same rotor.

A small amount of radioactive virus, labeled with [^{3}H]leucine, was added as a marker and the virus was further concentrated by pelleting onto a fluorocarbon cushion (Strauss *et al.*, 1969). In Spinco SW 27 tubes, 27.5 ml of virus solution was underlayered with 5.5 ml of 15% (w/v) sucrose in TNE, which was in turn underlayered with 4 ml Genesolv D (Allied Chemical, Morristown, N. J.). The virus was collected at the fluorocarbon–aqueous solution interface by centrifugation at 26,000 rpm for 2.5 hr at 4°. All but 2 to 3 ml of the supernatant was withdrawn and discarded and the virus was extracted into the aqueous phase by vigorous mixing with a Vortex mixer. After centrifugation at 12,000 rpm for 10 min at 4° in a Sorvall SS-34 rotor, the virus-containing aqueous layer was removed and the pellet-fluorocarbon layer extracted twice more with a small volume of TNE. The virus is almost 80% pure in terms of protein at this point (estimated from acrylamide gels of material stained with Coomassie brilliant blue) and has been concentrated about 200-fold.

The concentrated virus was further purified by rate zonal centrifugation through gradients of 15–30% (w/v) sucrose in TNE for 2 hr at 27,000 rpm and 4° in the Spinco SW 27 rotor. After fractionation the virus peak was located by assaying for radioactivity and further purified by isopycnic banding in linear sucrose-D_2O gradients (Strauss *et al.*, 1977). Light and heavy gradient solutions were 22 and 44% (w/v) sucrose, respectively. To a weighed amount of sucrose, 0.1 final vol of 10× TNE was added and the solutions were made up to volume with 99% D_2O. Densities of the light and heavy solutions (at 20°) were 1.17 and 1.24 g/ml, respectively. Pooled fractions from rate zonal centrifugation were layered (13–17 ml per tube) directly over 21–25 ml preformed sucrose-D_2O gradients in SW 27 tubes. After centrifugation at 24,000 rpm for 13.5 hr at 4°, fractions were collected by pumping from the bottom of the tube. Aliquots of each fraction were assayed for radioactivity and fractions containing the virus peak were pooled. The virus is essentially pure at this point.

In one large scale purification of HR virus, 38 mg of viral protein (assayed by the method of Lowry *et al.*, 1951) was obtained from the culture fluid of 375 infected roller bottles, and for the *wt* virus 18 mg of protein was obtained from 125 roller bottles. This represents a yield of about 60 to 70% of the starting material. The various steps of the purification procedure were developed because they allow progressive concentration as well as purification of the virus while no detectable change in the specific infectivity or physical properties of the virus particles occurs in any step.

Purification of the viral structural proteins. One-twelfth volume of 20% Triton X-100 in TN buffer (0.2 *M* NaCl, 0.05 *M* Tris, pH 7.4) was added to purified virus, previously dialyzed against TN buffer at 4°, and 8.3 ml of the resulting solution was underlayered with 4 ml of 20% sucrose, 0.05% Triton X-100 in TN buffer. After centrifugation at 41,000 rpm for 4 hr at 4° in an SW 41 rotor, the solution in the upper portion of the tube, corresponding to the original sample volume and containing E1 and E2, was removed. The remaining liquid was removed and discarded and the pelleted capsids were dissolved in 1% sodium dodecyl sulfate (SDS), 0.01 *M* Tris, pH 7.4, and 1% mercaptoethanol.

The capsid protein solution was made 0.20 *M* in $NaPO_4$, pH 6.4, and chromatographed on hydroxylapatite (Moss and Rosenblum,

1972) on a 1.8 × 2-cm column equilibrated with 0.20 *M* $NaPO_4$, pH 6.4, 0.1% SDS, 1 m*M* dithiothreitol (DTT). The capsid protein was eluted with a linear gradient of 0.20 to 0.50 *M* $NaPO_4$, pH 6.4, in 0.1% SDS, 1 m*M* DTT at a flow rate of 2.5 ml/min at 37°. Under these conditions the viral RNA is not bound to the column. The capsid protein was quantitatively recovered as a single peak eluting at 0.30 *M* phosphate.

The envelope proteins E1 and E2 were separated from each other by passage through a column of Pyrex glass wool (Corning Glass Works) which had been sheared to short fibers at high speed in a Waring Blendor. Before use, columns were washed with 1 liter of 1% SDS and equilibrated with 1 liter of 0.1% Triton X-100, 0.05 *M* Na succinate, pH 5.5, 0.3 *M* NaCl, 1 m*M* DTT. Samples, consisting of the upper portion of the gradient used to pellet the nucleocapsids, contained 0.3 mg protein/g glass wool at about 0.6 mg protein/ml. An equal volume of 0.1% Triton X-100, 0.1 *M* Na-succinate, pH 5.5, 0.4 *M* NaCl was added to the sample, which was applied to the column and washed through with the same buffer used for equilibration. The column was eluted with 0.1% Triton X-100, 0.05 *M* Tris-Cl, pH 8.0, 0.8 *M* NaCl, 1 m*M* DTT, and any material remaining bound to the column was recovered by further elution with 1% SDS, 0.05 *M* Tris–Cl, pH 8.0, 1 m*M* DTT. These are referred to as the Triton X-100 and SDS elutions, respectively. After pooling appropriate fractions, SDS was added to the E1 and E2 pools to a final concentration of 1%; the addition of SDS at this point was found to improve the recovery of the proteins during subsequent handling. Total recovery of protein from the column was greater than 80%. Cross-contamination was variable but less than 3–5%; in some cases, no cross-contamination could be detected (<0.5%) by slicing and counting analytical Laemmli gels run on the pooled fractions.

Solutions containing the three purified viral proteins were diluted with 0.05% SDS or distilled water and concentrated by ultrafiltration with an Amicon PM-10 membrane, diluted and reconcentrated so that the final volume was less than 15 ml and the salt concentration less than 0.05 *M*. Protein was precipitated overnight at −20° with 2 vol of ice-cold ethanol and collected by centrifugation at 15,000 rpm for 35 min at −20° in a Sorvall SS-34 rotor. Ethanol-precipitated proteins were dissolved in 0.5% SDS, 10 m*M* Tris, pH 7.5, 5 m*M* DTT, and stored frozen at −20°.

Overall recovery of protein, starting with purified virus, was on the order of 40–50%.

Amino acid analysis. Except as noted, all amino acids were determined after hydrolysis of duplicate samples with 6 *N* HCl at 110° for 24, 48, and 72 hr. Values obtained for serine and threonine were extrapolated to zero time of hydrolysis assuming first order kinetics. Isoleucine and valine were determined primarily from the samples hydrolyzed for 72 hr but data from 48 and 96 hr hydrolyses were also used. Cysteine was determined after HCl hydrolysis of performic acid-oxidized samples for 24 hr. Tryptophan was determined on single samples hydrolyzed for 15, 27, and 40 hr at 105° with mercaptoethanesulfonic acid. Values obtained were extrapolated to zero time of hydrolysis, assuming first-order kinetics (Penke *et al.*, 1974).

Automated amino acid analyses of protein hydrolysates were performed on the Durrum D-500 amino acid analyzer. For the determination of tryptophan, ethanol was added to the "C" buffer (1.0 *M* Na^+, pH 6.0) to a concentration of 10% (v/v). Under these conditions, tryptophan was eluted as a small broad peak just ahead of lysine.

Polyacrylamide gel electrophoresis. Quantitative gel electrophoresis was performed using the discontinuous buffer system of Laemmli (1970), except that the concentrations of Tris–chloride and Tris–glycine in the gels and samples were halved.

RESULTS

Purification of E1 and E2

While investigating the chromatographic behavior of E1 and E2 in Triton X-100, we noticed that E2 was bound to the glass wool used to plug the columns. We have exploited this binding as a means of separating E1 and E2, with typical results shown in Fig. 1. An analysis of pools of the peak fractions by SDS–gel electrophoresis is presented in

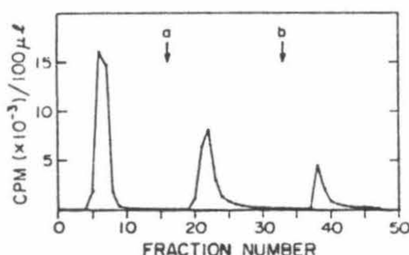


FIG. 1. Chromatography of E1 and E2 on glass wool. Protein (5.2 mg) (E1 and E2, prepared as described under Materials and Methods) was applied to the column, which was then washed with equilibration buffer. Buffer changes are indicated by arrows; the Triton X-100 elution (high salt, high pH) was begun at (a) and the SDS elution at (b). The column, 4.7 cm in diameter and containing 19 g of sheared glass wool, was operated at a flow rate of approximately 20 ml/min and 20-ml fractions were collected volumetrically.

Fig. 2. As can be seen from the figures, E1 is found in the column flow-through and wash, while E2 is recovered in the Triton X-100 and, to a lesser extent, in the SDS elutions. Cross-contamination is negligible.

The column used in the experiment shown was packed with glass wool which had been sheared to short fibers. When untreated glass wool was used, greater than 80% of the E2 was recovered in the SDS elution and little material was found in the Triton X-100 elution. Using sheared glass wool, however, the binding appears to be largely ionic, since much of the E2 is released from the glass support by a high salt, high pH buffer. The sites on the glass responsible for this binding are not known, but in preliminary experiments using unsheared glass wool, binding could be reduced by presoaking the glass in acid, base, or salt solutions. Extreme presoaking conditions, such as one normal acid or base overnight, resulted in a complete loss of binding. The columns bind a nearly constant amount of protein (E1 or E2) per gram of glass wool, so that underloading the column with respect to E2 results in a corresponding contamination of the E2 pool with E1.

The major difficulty in using the glass wool columns is the large sample dilution which sometimes results. Since it is difficult to pack glass wool so as to obtain an even buffer flow through the column, a variety of glasses such as Bio-Glas porous glass beads (Bio-Rad Laboratories) and several types of glass beads intended for use as a grinding media were tried as column material. None of the glass beads tested bound E2, although we have noticed that E2 is bound to soft glass Pasteur pipettes.

Amino acid analysis. The results of amino acid analyses of the structural proteins of the *wt* and HR strains of Sindbis virus are shown in Table 1, together with the standard errors calculated from the data. Also shown in the table are the calculated number of residues of each amino acid per molecule. The compositions of corresponding proteins of the two virus strains are quite similar, and it is difficult to attach

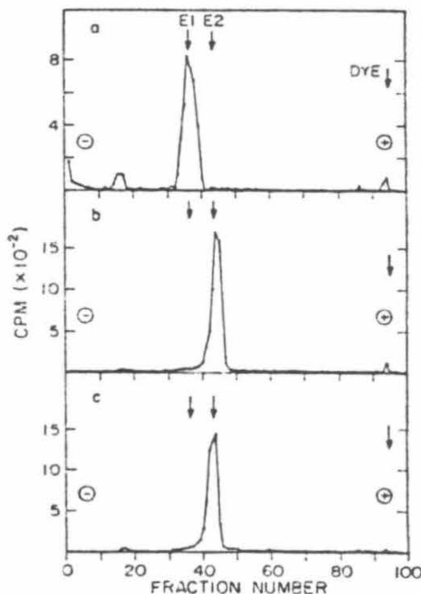


FIG. 2. Polyacrylamide gel analysis of pooled fractions from glass wool chromatography. Pooled fractions from the chromatography of E1 and E2 shown in Fig. 1 were analyzed by electrophoresis on Laemmli gels after precipitation with ethanol as described under Materials and Methods. After electrophoresis the stacking gel was discarded and the separating gel sliced into 1-mm fractions and counted in a scintillation fluid containing NCS tissue solubilizer. Electrophoresis was from left to right. Samples on individual gels were as follows: (a) 75 μl of the column flow-through and wash (fractions 6 and 7); (b) 180 μl of the Triton X-100 elution (fractions 21–23), (c) 400 μl of the SDS elution (fractions 38 and 39). The small peak near fraction 16 in panel a is an E1 dimer.

TABLE 1

AMINO ACID COMPOSITIONS OF THE STRUCTURAL PROTEINS FROM TWO STRAINS OF SINDBIS VIRUS[a]

	E1			E2			Capsid protein		
	wt	HR	No. of residues[b]	*wt*	HR	No. of residues[b]	*wt*	HR	No. of residues[b]
Aspartic acid + asparagine	7.73 (0.02)	7.79 (0.02)	35	7.96 (0.02)	7.74 (0.02)	35	6.41 (0.01)	6.33 (0.01)	17
Threonine	7.82 (0.09)	7.77 (0.08)	35	9.29 (0.11)	9.35 (0.05)	42	6.54 (0.05)	6.53 (0.06)	18
Serine	9.20 (0.12)	9.78 (0.08)	43	6.84 (0.17)	7.18 (0.08)	32	4.61 (0.02)	4.56 (0.07)	12
Glutamic acid + glutamine	7.88 (0.01)	8.02 (0.01)	36	6.78 (0.05)	7.07 (0.01)	31	9.86 (0.03)	10.04 (0.01)	27
Proline	6.57 (0.04)	6.35 (0.06)	29	6.65 (0.12)	7.06 (0.04)	31	10.35 (0.13)	10.80 (0.04)	28
Glycine	6.17 (0.01)	6.29 (0.01)	28	6.38 (0.03)	6.28 (0.01)	28	9.15 (0.02)	9.06 (0.01)	24
Alanine	9.65 (0.01)	9.68 (0.01)	43	7.85 (0.01)	7.82 (0.01)	35	8.39 (0.01)	8.38 (0.01)	22
Cysteine[c]	4.31	3.50	18	3.46	3.26	15	N.F.[d]	N.F.[d]	0
Valine	7.89 (0.02)	8.01 (0.03)	36	7.77 (0.03)	7.40 (0.01)	34	5.97 (0.01)	5.81 (0.01)	16
Methionine	2.15 (0.02)	2.22 (0.04)	10	1.54 (0.01)	1.49 (0.01)	7	3.67 (0.06)	3.69 (0.02)	10
Isoleucine	5.23 (0.03)	5.31 (0.01)	24	5.35 (0.02)	5.27 (0.04)	24	3.06 (0.01)	3.00 (0.01)	8
Leucine	6.28 (0.03)	6.38 (0.03)	28	6.96 (0.02)	6.92 (0.01)	31	5.54 (0.02)	5.47 (0.01)	15
Tyrosine	3.05 (0.06)	2.97 (0.07)	14	4.31 (0.05)	4.20 (0.05)	19	1.47 (0.01)	1.47 (0.02)	4
Phenylalanine	4.18 (0.02)	4.44 (0.01)	19	2.46 (0.03)	2.36 (0.01)	11	3.56 (0.02)	3.48 (0.01)	9
Histidine	3.03 (0.01)	2.91 (0.01)	13	4.22 (0.01)	4.12 (0.01)	19	2.23 (0.02)	2.25 (0.01)	6
Lysine	5.30 (0.01)	5.26 (0.01)	24	6.37 (0.01)	6.46 (0.02)	29	9.39 (0.02)	9.42 (0.02)	25
Arginine	2.42 (0.01)	2.31 (0.01)	11	4.22 (0.01)	4.57 (0.01)	20	8.14 (0.04)	8.28 (0.03)	22
Tryptophan	1.15 (0.07)	1.02 (0.08)	5	1.60 (0.12)	1.46 (0.22)	7	1.65 (0.03)	1.44 (0.05)	4

[a] Expressed as mole percentages. Standard errors are in parentheses.

[b] Number of residues of each amino acid per molecule, calculated from the means of the *wt* and HR mole percentages and assuming a total of 450 residues each for E1 and E2 and a total of 270 residues for the capsid protein. The results are rounded to the nearest integer.

[c] Determined as cysteic acid.

[d] None found (<0.15).

any significance to those differences which are found. In preliminary determinations on an independent preparation of the HR strain, relative amounts of the amino acids were found which, while in good agreement with the HR results in Table 1, did sometimes differ by somewhat more than would be expected on the basis of statistical considerations (results not shown). We note that the capsid protein contains a relatively large proportion of proline and of the basic amino acids lysine and arginine. No cysteine could be detected in the capsid protein. The polarity indexes of the proteins, a measure of the relative amounts of amino acids with polar side chains as proposed by Capaldi and Vanderkovi (1972), are 44, 46, and 47 for E1, E2, and the capsid protein, respectively, similar to values typically found for water soluble proteins.

Estimation of the Molecular Weight of the Capsid Protein

A value for the molecular weight of the capsid protein can be calculated from the composition data by choosing an integral number of residues per molecule for the least commonly occurring amino acids. Since the data for the two least common amino acids, tryptophan and tyrosine, are among the less reliable, we have chosen to consider histidine and isoleucine in addition to these two. Using the data for the HR strain, we obtain the values of 278 total residues, assuming four tryptophans, 272 total residues from four tyrosines, 267 total residues from six histidines, and 267 total residues from eight isoleucines. The next best choices, about 200 total residues (from three tryptophans, three tyrosines, and six isoleucines) or about 340 total residues (from five tryptophans, five tyrosines, and ten isoleucines) are both inconsistent with an integral number of histidines. Either would also imply a 25% error in the gel molecular weight of 30,000–34,000 daltons (Strauss *et al.*, 1969; Strauss, 1978). From a total of 270 ± 5 amino acid residues, we calculate the molecular weight of the capsid protein to be $30{,}000 \pm 600$ daltons.

DISCUSSION

The purification of the envelope glycoproteins E1 and E2 of Sindbis virus, which requires the use of detergents for solubilization, presents special problems. When solubilized with SDS, the size difference is too minor to be used as a convenient basis for separation, and when Triton X-100 is used instead of SDS, E2 precipitates unless a relatively high ionic strength is maintained (Burke and Keegstra, 1976; J. R. Bell, unpublished observations), although E1 and E2 have been purified by isoelectric focusing on sucrose gradients containing Triton X-100 (Dalrymple *et al.*, 1976). However, under appropriate conditions, E2 shows much greater binding to certain types of glass than does E1. We have found that chromatography on columns packed with glass wool can thus be used to prepare milligram amounts of pure preparations of E1 and E2. As this purification is performed in the presence of the nonionic detergent Triton X-100, one can reasonably hope that much of the native conformation and biological activity of these proteins is preserved (Helenius and Simons, 1975).

We have determined the amino acid compositions of the structural proteins of two strains of Sindbis virus, the *wt* and the HR strains. Because the HR strain was derived from the *wt* by multiple cycles of selection for thermal stability of the virion (Burge and Pfefferkorn, 1966), there must be some differences between the structural proteins of the two strains. Our data show that these differences are so minor that they cannot be reliably detected at the level of amino acid composition of the proteins. Thus we are unable to determine at this point which protein or proteins have been altered to produce this phenotypic change. The amino acid compositions of the Sindbis proteins have been reported previously by other workers (Burke and Keegstra, 1976). Differences exist between the amino acid compositions reported in this previous work and the results presented here.

The Sindbis glycoproteins E1 and E2 are known to be integral membrane proteins and the hydrophobic interactions between the proteins and the lipids of the envelopes are believed to arise from the presence of regions of the polypeptide chains containing almost exclusively hydrophobic amino acids. Our data indicate that these hydrophobic regions must be relatively short, since there is no apparent excess of hydrophobic amino acids in the glycoproteins as a whole. This is in agreement with other experiments in which the hydrophobic regions of the envelope glycoproteins, isolated by digesting those parts of the proteins external to the lipid bilayer with proteolytic enzymes, were found to be on the order of 5000–10,000 daltons (Garoff and Söderlund, 1978; C. M. Rice, personal communication).

In comparison with the glycoproteins, the amino acid composition of the capsid protein clearly reflects the function of the protein. Because of its association with the RNA of the virus, the protein contains a high content of amino acids with basic side chains. Together, lysine, histidine, and arginine account for one-fifth of the amino acids of the capsid protein. It is interesting to note the degree of similarity between the capsid proteins of Sindbis and Semliki Forest viruses in this respect. These two viruses have been shown to be closely related on the basis of antigenic cross-reactivity (Casals and Brown, 1954), as well as by the composition and structure of the virion and the biochemical events which occur during virus replication (Strauss and Strauss, 1977). We calculate that a molecule of the Sindbis capsid protein contains 22 arginine residues, 25 lysines, and six histidines. From the amino acid composition data of Kennedy and Burke (1972) or Simons and Kääriäinen (1970) we calculate that the Semliki Forest virus capsid protein contains eight or nine fewer arginine residues, but nine more lysines and one more histidine per molecule than Sindbis, and it therefore appears there has been a conservative replacement of numerous arginines by lysines. It is of interest to note that the Sindbis capsid protein contains no cysteine, while it has been found that cysteine constitutes 1.5–1.6% of the amino acids of the Semliki Forest virus capsid protein, or four cysteine residues per molecule (Kennedy and Burke, 1972; Simons and Kääriäinen, 1970).

Although we have obtained a molecular

weight estimate for the capsid protein from the composition data, this approach could not be applied with any confidence to E1 and E2 due to the lack of any amino acids present in sufficiently low amounts. We have assumed a molecular weight of 53,000 daltons for both E1 and E2 based on their migration, as a single band, in SDS–gel electrophoresis at neutral pH (Strauss *et al.*, 1969). After subtracting 0.08 mg carbohydrate/mg protein (Sefton and Keegstra, 1974), we are left with a polypeptide molecular weight of 49,000, or about 450 amino acids per molecule. However, the tendency to overestimate the molecular weights of membrane glycoproteins in SDS–gel electrophoresis is well known. The effect of the small amount of carbohydrate in E1 and E2 is likely to be slight, for the use of tunicamycin, a drug which inhibits glycosylation of proteins, produces an altered E1 which, from the data of Leavitt *et al.* (1977), migrates on gel electrophoresis as a protein 10% smaller than the glycosylated form. A more serious cause of reservation is the fact that the glycoproteins of the closely related Semliki Forest virus bind an anomalously large amount of SDS (Simons and Kääriäinen, 1970), a factor likely to affect their migration rate during electrophoresis. In addition, E1 and E2 show an apparent difference of 6000 daltons on electrophoresis in the Laemmli gel system (Strauss, 1978), although we have recently noted (C. M. Rice and J. R. Bell, unpublished observations) that this difference is abolished when various "electrophoresis purity" SDS preparations are used. Matheka *et al.* (1977), using a different gel system, have also reported that the source of the SDS used affects the migration rate of Sindbis proteins and of the proteins of foot-and-mouth disease virus, and presented data that such changes in migration rates may be due to the presence or absence of hydrocarbon chains greater than 12 carbons in length. The accurate determination of the polypeptide molecular weights of E1 and E2 clearly awaits further investigation.

ACKNOWLEDGMENTS

We thank Edith Lenches and Mary Stammreich-Martin for their assistance in growing the virus used in these studies. We also wish to thank Dr. Leroy E. Hood and members of his research group for making available to us their facilities and expertise for the determination of the amino acid compositions; this collaboration was supported by Grant GM 06965 from the U. S. Public Health Service. This work was also supported by Grant AI 10793 from the U. S. Public Health Service and Grant PCM 77-26728 from the National Science Foundation.

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. Nat. Acad. Sci. USA* 75, 2722–2726.

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.

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

BURKE, D. J., and KEEGSTRA, K. (1976). Purification and composition of the proteins from Sindbis virus grown in chick and BHK cells. *J. Virol.* 20, 676–686.

CAPALDI, R. A., and VANDERKOVI, G. (1972). The low polarity of many membrane proteins. *Proc. Nat. Acad. Sci. USA* 69, 930–932.

CASALS, J., and BROWN, L. V. (1954). Hemagglutination with arthropod-borne viruses. *J. Exp. Med.* 99, 429–449.

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

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.

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., 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.

HELENIUS, A., and SIMONS, K. (1975). Solubilization of membranes by detergents. *Biochim. Biophys. Acta* 415, 29–79.

KENNEDY, S. I. T., and BURKE, D. C. (1972). Studies on the structural proteins of Semliki Forest virus. *J. Gen. Virol.* 14, 87–98.

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

LEAVITT, R., SCHLESINGER, S., and KORNFELD, S. (1977). Tunicamycin inhibits glycosylation and multiplication of Sindbis and Vesicular Stomatitis viruses. *J. Virol.* 21, 375–385.

LOWRY, O. H., ROSEBROUGH, N. J., FARR, A. L., and RANDALL, R. J. (1951). Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* 193, 265–275.

MATHEKA, H. D., ENZMANN, P.-J., BACHRACH, H. L., and MIGL, B. (1977). The influences of sodium dodecyl sulfate from different sources on the separation of virus proteins in polyacrylamide gel electrophoresis. *Anal. Biochem.* 81, 9–17.

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

PENKE, B., FERENCJI, R., and KOVACS, K. (1974). A new acid hydrolysis method for determining tryptophan in peptides and proteins. *Anal. Biochem.* 60, 45–50.

PFEFFERKORN, E. R., and CLIFFORD, R. L. (1964). The origin of the proteins of Sindbis virus. *Virology* 23, 217–223.

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.

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

SEFTON, B. M., and KEEGSTRA, K. (1974). Glycoproteins of Sindbis virus: preliminary characterization of the oligosaccharides. *J. Virol.* 4, 522–530.

SIMONS, K., and KÄÄRIÄINEN, L. (1970). Characterization of the Semliki Forest virus core and envelope proteins. *Biochem. Biophys. Res. Commun.* 38, 981–988.

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., BIRDWELL, C. R., LENCHES, E. M., STAPLES, S. E., and STRAUSS, J. H. (1977). Mutants of Sindbis virus. II. Characterization of a maturation-defective mutant, ts103. *Virology* 82, 122–149.

STRAUSS, J. H., BURGE, B. W., and DARNELL, J. E. (1969). Sindbis virus infection of chick and hamster cells: synthesis of virus-specific proteins. *Virology* 37, 367–376.

STRAUSS, J. H., BURGE, B. W., and DARNELL, J. E. (1970). Carbohydrate content of the membrane protein of Sindbis virus. *J. Mol. Biol.* 47, 437–448.

STRAUSS, J. H., BURGE, B. W., PFEFFERKORN, E. R., and DARNELL, J. E. (1968). Identification of the membrane and "core" protein of Sindbis virus. *Proc. Nat. Acad. Sci. USA* 59, 533–537.

STRAUSS, J. H., and STRAUSS, E. G. (1977). Togaviruses. *In* "The Molecular Biology of Animal Viruses" (D. Nayak, ed.), Vol. 1, pp. 111–166. Dekker, New York.

WIRTH, D. F., KATZ, F., SMALL, B., and LODISH, H. F. (1977). How a single Sindbis virus mRNA directs the synthesis of one soluble protein and two integral membrane glycoproteins. *Cell* 10, 253–263.

Chapter 3

Amino-terminal Sequence Analysis of the Structural Proteins of Sindbis Virus

John R. Bell, Michael W. Hunkapiller, Leroy E. Hood, and James H. Strauss

This chapter was published in PNAS

Reprinted from

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

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

(membrane glycoproteins/togaviruses/microsequencing)

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

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

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

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

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

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

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

MATERIALS AND METHODS

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

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

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

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

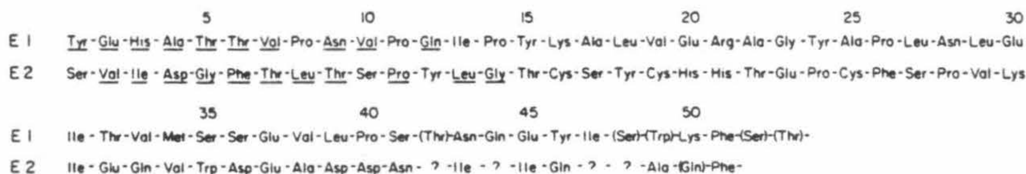


FIG. 1. NH_2-Terminal amino acid sequences of the Sindbis virus envelope proteins, E1 and E2. ( ), Some uncertainty in assignment; ?, no assignment. Underlined amino acids were also identified in preliminary runs on the solid-phase sequenator.

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

RESULTS

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

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

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

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

DISCUSSION

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

Table 1. Composition of Sindbis glycoproteins

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

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

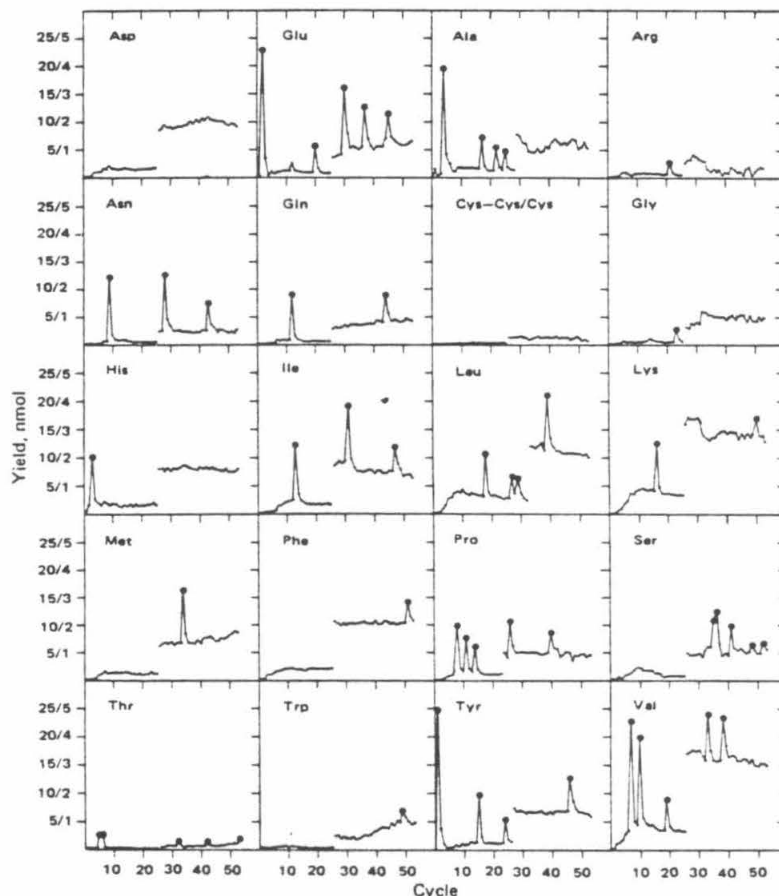


FIG. 2. Yields of amino acid phenylthiohydantoins from an NH_2-terminal spinning cup sequenator analysis of 50 nmol of Sindbis glycoprotein E1 (peptide molecular weight, ~45,000). Aliquots of each cycle were analyzed by high-pressure liquid chromatography, peaks were quantitated by comparison with a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

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

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

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

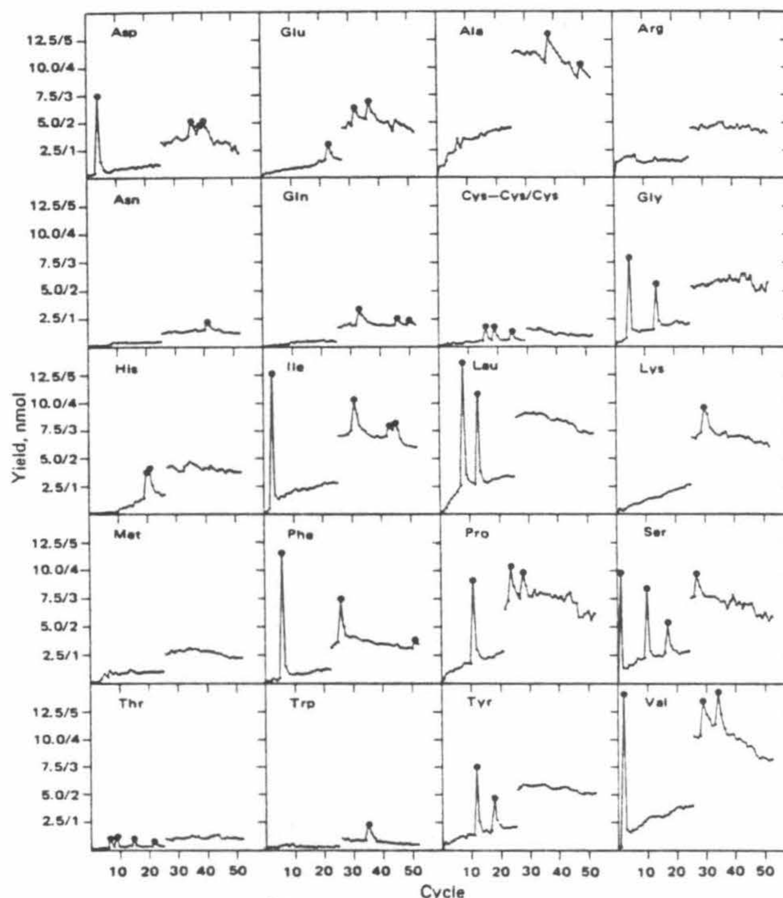


FIG. 3. Yields of amino acid phenylthiohydantoins from an NH_2-terminal spinning cup sequenator analysis of 25 nmol of Sindbis glycoprotein E2 (peptide molecular weight, ~48,000). Yields were calculated as described in the legend to Fig. 2, and the data are plotted in a similar fashion with a 2.5-fold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

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

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

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

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

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

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

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

Chapter 4

In Vivo N-terminal Acetylation of Sindbis Virus Proteins

JOHN R. BELL and JAMES H. STRAUSS

This chapter has been prepared for submission
to the Journal of Biological Chemistry

ABSTRACT

The *in vivo* incorporation of exogenous radioactive acetate into two proteins of Sindbis virus, the capsid protein and PE2, is described. Under appropriate labeling conditions, 40-50% of the label in the capsid protein is found in an N-acetyl group which constitutes the N-terminal modification of this blocked protein. The incorporated radiolabeled acetate was useful in the purification and analysis of peptides derived from the N-terminus of the capsid protein, and from these peptides the N-terminal sequence of the protein was determined to be N-acetyl-met-asx-, with the asx group most likely asparagine. The analysis of a peptide derived from the N-terminus of PE2 and containing 45% of the acetate-derived label in this protein leads us to conclude that at least a significant fraction of PE2 is also blocked by N-acetylation.

INTRODUCTION

The virion of Sindbis virus (a member of the alphavirus group of the Togavirus family) is composed of three proteins, the capsid protein (C), mw 30,000 daltons, and the envelope glycoproteins E1 and E2, each with a polypeptide mw about 50,000 daltons. E2 is found in the host cell in the form of a precursor, PE2, and the cleavage of this precursor occurs rather late in the maturation of the virus particle. The N-terminal portion of PE2 is of particular interest, since it contains the sequences necessary for the transmembrane translocation of this protein during synthesis (Garoff et al., 1978). The structure of the virus and the synthesis of the viral proteins have been reviewed recently (Strauss and Strauss, 1977).

Previously during the course of our studies on the primary sequence of the viral structural proteins, we found that the capsid protein contains a modified, or blocked, N-terminus and so could not be analyzed by Edman degradation (Bell et al., 1978). The N-terminal modification in blocked proteins often consists of the addition of an acetyl group (Brown and Roberts, 1976), and this is the only N-terminal modification found in the blocked capsid proteins of viruses (see Narita et al., 1975). We therefore looked for an N-acetyl group in the Sindbis capsid protein by analyzing the in vivo incorporation of radioactive acetate into this protein. This approach was suggested to us by the work of Jörnvall et al. (1974), who incorporated ^{14}C-acetate into the acetylated N-terminus of the hexon protein of adenovirus. These authors noted complications due to the metabolic conversion of the label into other chemical forms (Jörnvall et al., 1978), which we also observed in preliminary experiments.

In this paper we describe the in vivo incorporation of tritiated acetate into the blocked N-terminus of the capsid protein of Sindbis virus with improved specificity, and report the primary sequence of its N-terminus. In the course of this study, PE2 was also found to be specifically labeled with acetate. This unexpected result led us to examine this protein in more detail, and we show that a significant fraction of PE2

also has an acetylated N-terminus. The results of this study may be of use in the identification of other N-acetylated proteins and the determination of their N-termini.

MATERIALS AND METHODS

Virus Strains and Cell Infection

The virus strain used in these studies was the large plaque variant of the HR strain of Sindbis virus, originally obtained from Dr. B. W. Burge and plaque purified several times in our laboratory. The virus was grown in chick embryo fibroblasts (primary cultures, except as noted below) or BHK-21 cells (the generous gift of Dr. B. M. Sefton). The preparation of chick cells and infection with Sindbis virus were as described (Pierce *et al.*, 1974).

Growth and Purification of Viral Proteins Radiolabeled with Amino Acids

When viral proteins labeled with an essential amino acid were to be prepared, the medium used for incubation post-infection and for labeling contained one-tenth the normal concentration of that amino acid, except as noted. At the end of the labeling period, usually from 7 to 8 hr post-infection, monolayers were washed twice with ice-cold phosphate buffered saline and dissolved in a small volume of 1% sodium dodecyl sulfate (SDS), 10 m*M* Tris-Cl, pH 7.5.

Viral proteins were purified from infected cell monlayers, or from the virus released from infected cells by a high salt wash (Pierce *et al.*, 1974), by preparative polyacrylamide gel electrophoresis. Protein was precipitated by the addition of two volumes of ethanol, frozen on dry ice, warmed to 0°C, and centrifuged for 15 min at 15,000 RPM at 0° in a Sorvall SS-34 rotor. Precipitated protein was dissolved in sample buffer and electrophoresed in 8 mm diameter polyacrylamide gels using the buffer system of Laemmli (1970), except that the concentration of Tris in the gels

was halved and 0.01% mercaptopropionic acid was included in the sample and upper electrode buffer. After electrophoresis, the gel was sliced into approximately 100 fractions, each consisting of two 0.5 mm slices, and each fraction was eluted in 1 ml of 0.5% SDS, 10 m$\underline{M}$ 2-(N-morpholino)ethanesulfonic acid (MES), 2.9 m$\underline{M}$ NaOH, 1 m$\underline{M}$ dithiothreitol, 20 μg/ml phenylmethylsulfonyl fluoride. After shaking for 9 hr at room temperature, an aliquot of each fraction was assayed for radioactivity by liquid scintillation counting. Appropriate fractions were pooled and the volume of each pool was reduced ten-fold by lyophilization. Just before use, purified proteins were precipitated with ethanol as described above with tuna heart cytochrome $\underline{c}$ added as carrier to at least 20 μg/ml.

Growth and Purification of Viral Proteins Radiolabeled with Tritiated Acetate

Two days before use, cell monolayers were harvested by trypsinization and reseeded in the following medium. To Eagle's minimal essential medium (Eagle, 1959) without NaCl was added L-aspartic acid to 2 m$\underline{M}$ (from a 0.1 $\underline{M}$ stock in 0.2 $\underline{M}$ HCl), L-glutamic acid·HCl to 15 m$\underline{M}$, L-proline to 30 m$\underline{M}$, NaCl to 33 m$\underline{M}$, and NaOH to 36 m$\underline{M}$. The amounts of NaCl and NaOH added were calculated to preserve neutrality and isotonicity. Fetal calf serum was added to this mixture to a concentration of 10% (v/v). Chick embryo fibroblasts did not grow well in this medium and so were not split during the transfer, while BHK-21 cells were split two-fold. Virus infection and post-infection incubation were carried out in the above medium with 10% dialyzed fetal calf serum in place of fetal calf serum and containing 1 μg/ml actinomycin D. At about 7 hr postinfection, cells were labeled for 20 min by the inclusion of 1 mCi/ml ^{3}H-acetate (2 Ci/mmole) in the medium and harvested as described above.

Preparative polyacrylamide gel electrophoresis of acetate labeled proteins was preceded by an additional purification step of affinity chromatography which will be

described in detail elsewhere (Chapter 5, this thesis). Briefly, monolayers harvested with SDS were diluted with a buffer containing Triton X-100 and passed over a column containing covalently coupled antibodies (anti-C and anti-E2) produced in rabbits in response to purified SDS-denatured proteins. (Anti-E2 also binds PE2.) Bound protein was eluted with SDS and purified by preparative polyacrylamide gel electrophoresis as above.

Staphylococcal Protease Digestion

Ethanol precipitated radiolabeled capsid protein plus cytochrome c carrier was resuspended in 67 µl of 0.1% SDS, 0.05 M $NaPO_4$, pH 7.8, 1 mM ethylenediamine tetraacetic acid containing Staphylococcus aureus V8 protease to give a carrier protein concentration of 0.8 mg/ml and a protease concentration of 0.1 mg/ml. In this buffer, the protease cleaves peptide bonds involving the carboxyl groups of aspartic acid and glutamic acid (Drapeau et al., 1972). After incubation for 4 hr at 37°, the sample was lyophilized and analyzed by electrophoresis on cylindrical polyacrylamide gels containing SDS, urea, and a high concentration of acrylamide (Swank and Munkres, 1971). After electrophoresis, the gel was sliced into 1 mm fractions and counted by liquid scintillation. Identical results were found when the digestion was extended an additional 4 hr.

Chymotrypsin Digestion and Dowex 50 W Chromotography

Ethanol precipitated radiolabeled capsid protein or PE2 was suspended in 0.05 M NH_4HCO_3, pH 7.8, usually 0.5 ml, to give a protein concentration (carrier or chemically pure capsid protein) of 120 µg/ml. One-tenth volume of TPCK-treated chymotrypsin at 2 mg/ml was added, and the mixture was incubated for 12 hr at room temperature. An additional one-tenth volume of chymotrypsin at 2 mg/ml was added and the incubation continued an additional 12 hr, at which time the sample was lyophilized, dissolved in a small volume of distilled water, and lyophilized again. The

sample was dissolved in 200 μl of 0.5% β-mercaptoethanol and applied to a 0.7 x 4 cm column of AG50W-X2 (H^+form), 200-400 mesh (purified Dowex resin from Bio-Rad Laboratories), in distilled water. The column was run at a flow rate of 1.35 ml/hr at 4° and 1-min fractions were collected. The column was washed with distilled water for 8 min, and then bound peptides were eluted with 1 M NaOH. Fractions containing base were neutralized with HCl and an aliquot of each fraction was assayed for radioactivity by liquid scintillation counting.

Chymotrypsin digestion and Dowex 50W chromatography were performed at room temperature and 4°, respectively, to minimize cyclization of newly exposed N-terminal glutamine or glutamic acid residues to pyrolidone carboxylyl groups (Ikenaka et al., 1966).

Analysis of Dowex 50W Unbound Peptides

Pooled fractions from Dowex 50W chromatography were dried under nitrogen at 50°, dissolved in 10 μl of 1% β-mercaptoethanol, and analyzed in one of two ways. Silica gel chromatography utilized ITLC SA "instant thin layer chromatography" sheets (Gelman Instrument Company). Ascending chromatography was performed with the solvent n-butanol-acetic acid-water (4:1:1). The dried chromatogram was cut into 3 mm fractions which were counted directly in Aquasol II (New England Nuclear) containing 6% water. Alternatively, samples were analyzed by reversed phase high pressure liquid chromatography (HPLC) on 4.6 mm x 25 cm columns (Dupont Instruments) at 1 ml/min at room temperature. Elution was with a gradient from A buffer (0.1% triflouroacetic acid) to B buffer (0.1% trifluoroacetic acid, 33% methanol, 33% acetonitrile, 33% water). For peptides derived from PE2, a Zorbax CN column was used, and a 30 min linear gradient from A buffer to 25% B buffer (v/v in A buffer) was applied 10 min after injection, followed by a 30 min linear gradient to 100% B buffer and continued elution with B buffer. Fractions of 1 ml were collected and an aliquot of each fraction was analyzed for radioactivity. A Zorbax

ODS column was used for the analysis of peptides derived from the capsid protein, and in this case a 30 min linear gradient from A buffer to 20% B buffer was applied starting at 10 min after injection, and 0.7 min fractions were collected. Pooled fractions from HPLC were dried under nitrogen at 50°, hydrolyzed in evacuated tubes with 6 N HCl at 145° for 4 hr, and analyzed on a Durrum D-500 mk. II amino acid analyzer. Chromatographic markers were prepared by acetylation of ^{35}S-methionine, containing a small amount of ^{35}S-methionine sulfoxide, with acetic anhydride in glacial acetic acid (Kolb and Toennies, 1942).

Miscellaneous

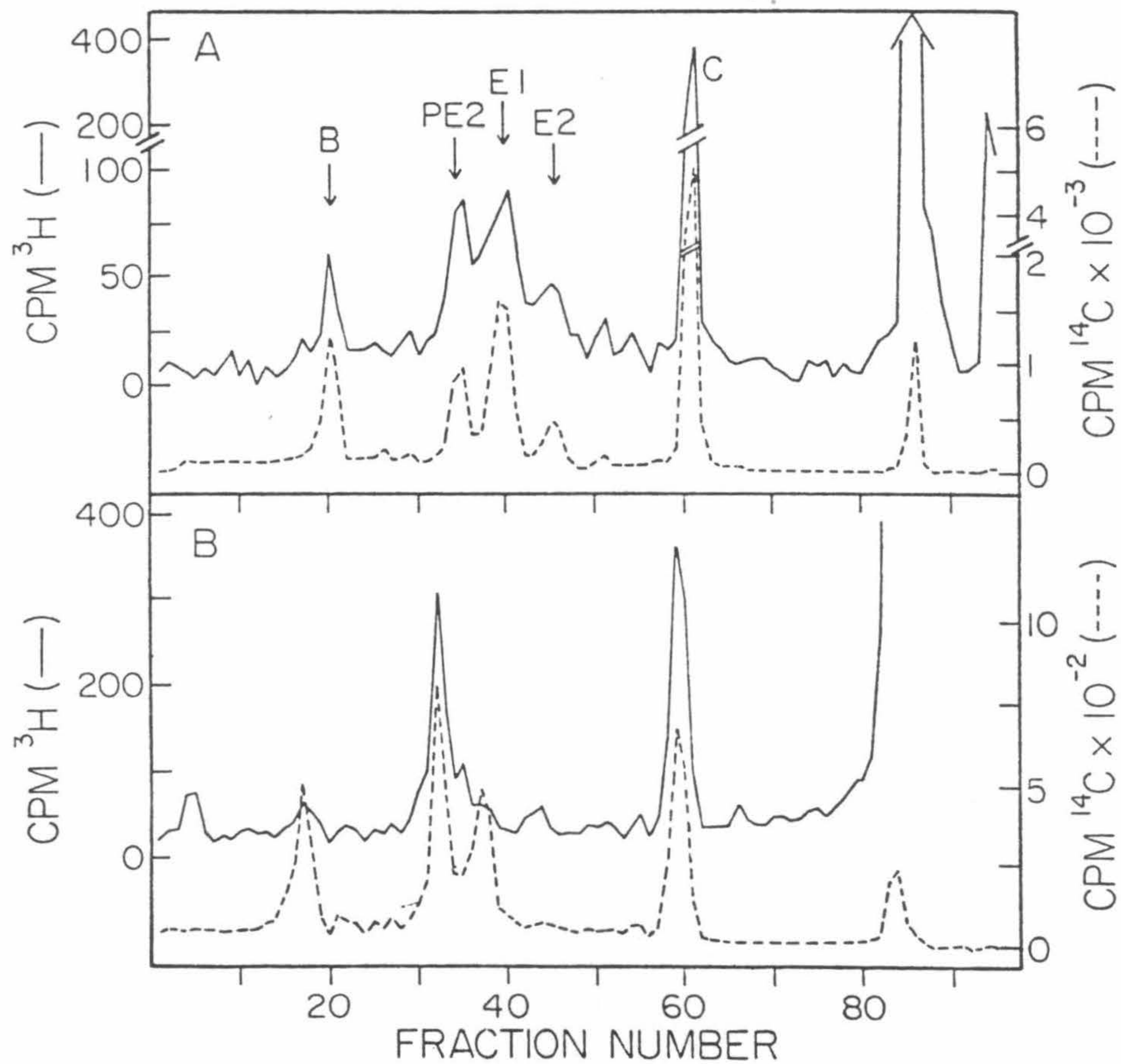
The preparation of chemically pure capsid protein, free of other proteins and RNA (Bell *et al.*, 1979), and of chemically pure PE2 (Chapter 5, this thesis) are described elsewhere. In all double label experiments, the results were corrected for channel overlap with the use of appropriate standards.

RESULTS

Radiolabeling with ^{3}H-acetate

In preliminary experiments, when infected cells were incubated with tritiated acetate, harvested, and analyzed by polyacrylamide gel electrophoresis, it was found that E1 (which is not blocked in the virion [Bell *et al.*, 1978] and is not known to be otherwise acetylated) contained a significant amount of tritium label (Fig. 1a). Since acetate occupies a central position in the intermediary metabolism of the host cell, one would expect to find the radioactivity derived from exogenous acetate in a variety of chemical forms including long chain fatty acids and the amino acids synthesized from citric acid cycle intermediates. Therefore, viral proteins were purified from infected cells which had been labeled with ^{3}H-acetate in a common growth medium (Eagle's minimal essential medium [Eagle, 1959] plus 2% fetal calf

FIG. 1. Acetate labeled Sindbis infected cells. Infected BHK-21 monolayers were labeled with tritiated acetate at 1 mCi/ml (2 Ci/mmole), harvested as described in Materials and Methods, and an aliquot was analyzed by polyacrylamide gel electrophoresis with the buffer system of Laemmli (1970). After electrophoresis, the gel was sliced into 1 mm fractions and analyzed for radioactivity by liquid scintillation counting. Electrophoresis was from left to right, and the positions of the viral proteins are identified in the figure. (a) The cells were labeled for 1 hr in minimal essential medium (Eagle, 1959) containing one-fortieth the usual concentration of methionine, 2% dialyzed fetal calf serum, 1 μg/ml actinomycin D, and also 3 μCi/ml ^{35}S-methionine (1.2 Ci/mmole final). The dye marker (bromophenol blue) was located at fraction 87. Note the corresponding breaks in the tritium and ^{14}C scales in the capsid protein peak. (b) The cells were labeled as described for acetate labeling in Materials and Methods with medium containing, in addition, 0.17 μCi/ml ^{14}C-leucine (0.4 mCi/mmole final). The dye marker was located at fraction 85.


A
B
PE2
E1
E2
C
CPM ^{3}H (—)
CPM ^{14}C × 10^{-3} (----)
CPM ^{14}C × 10^{-2} (----)
FRACTION NUMBER
400
200
100
50
0
300
6
4
2
1
10
5
20
40
60
80

serum) and the proteins were analyzed to determine the fate of the radioactive metabolite. After HCl hydrolysis, greater than 60% of the tritium label was found in the amino acids proline (about three-fourths of the total amino acid label), glutamic acid, and, to a minor extent, aspartic acid (data not shown). The Sindbis glycoproteins also contain covalently coupled long chain fatty acids (Schmidt et al., 1979). These fatty acids were released from the purified proteins after base hydrolysis in methanol, extracted with organic solvents, and the O-methyl esters of long chain fatty acids were identified by adsorption chromatography on silica gel (results not shown). Radioactivity was found in these long chain fatty acids under the above labeling conditions. Surprisingly, no radioactivity was found in the glycopeptides, which contain N-acetyl glucosamine and sialic acid, generated from the Sindbis glycoproteins by pronase digestion and analyzed as described (Sefton and Keegstra, 1974).

In view of these findings, in all subsequent experiments we used a modified labeling procedure, as described in Materials and Methods, which consisted of preincubation of the cells as well as labeling in a medium containing higher than normal concentrations of serum and large amounts of aspartic acid, proline, and glutamic acid. It was also necessary to use short labeling periods, and we found 20 min to be optimum (data not shown). The effects of this modification of the labeling procedure can be seen in Fig. 1, which shows the SDS polyacrylamide gel patterns of BHK-21 cells labeled with ^{3}H-acetate and ^{35}S-methionine or ^{14}C-leucine. The results in Fig. 1a were obtained during preliminary experiments with no attempt to suppress metabolic conversion of acetate, while the infected cells of Fig. 1b were labeled under the modified labeling protocol. The ratio of tritium to ^{35}S-methionine is similar in all of the structural proteins of the virus in Fig. 1a, but in Fig. 1b E1 is not extensively labeled with tritium while the capsid protein has still incorporated significant amounts of the label. Interestingly, even with the modified

labeling procedure, PE2 is labeled with tritium to nearly the same extent as the capsid protein. The amount of tritium incorporated into either PE2 or the capsid protein is quite low and there is a very large tritium peak near the dye front.

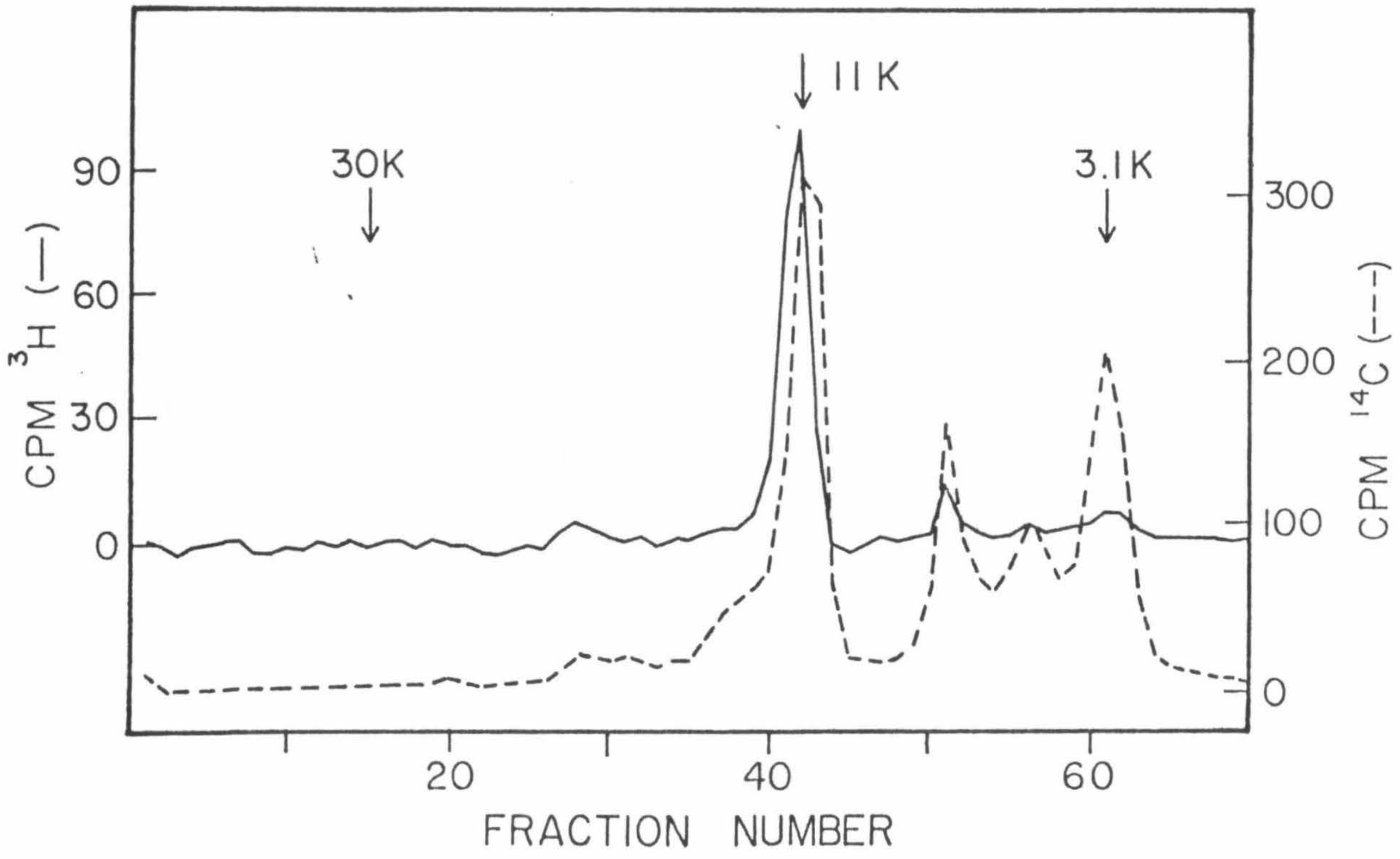
A similar result is found when infected chick cells are labeled in this manner. The ratio of tritium to ^{14}C in PE2 and in the capsid protein is similar to that seen in BHK-21 cells when labeling medium containing the same ratio of ^{3}H-acetate and ^{14}C-leucine is used. However, in chick cells a large unidentified tritium peak, which we have also observed in uninfected cells, obscures the E1 peak and seems to be more prominent in the gel patterns of short labeling periods. A small amount of this unidentified host-derived material can also be seen in BHK-21 cells (fraction 35 in Fig. 1b).

In order to examine the distribution of acetate-derived tritium label along the polypeptide chain of the capsid protein, capsid protein from chick cells was digested with S. aureus V8 protease and the digest was analyzed by SDS-polyacrylamide gel electrophoresis. As shown in Fig. 2, about 70% of the tritium label was found in one polypeptide fragment, which was also labeled extensively with methionine. The acetate and ^{14}C labels in this experiment were derived from separate preparations of the capsid protein, and the major peaks migrate at slightly different rates because one or more methionine residues in the acetate-labeled protein became oxidized during purification. A similar difference in migration rate during SDS-gel electrophoresis due to oxidation is also seen in the intact capsid protein (C. M. Rice, personal communication).

Acetate labeled peptides of the capsid protein

Having developed conditions for specifically labeling the blocked capsid protein with acetate and having shown that the label was incorporated in this protein in a non-random manner, we next purified small acetate-labeled peptides from proteolytic digests of the capsid protein. We used chromatography of an exhaustive chymotrypsin

FIG. 2. *S. aureus* V8 peptides of acetate labeled capsid protein. The peptides derived from chick embryo fibroblast grown capsid protein, labeled with tritiated acetate and ^{14}C-methionine, were analyzed by SDS polyacrylamide gel electrophoresis. Electrophoresis was from left to right. Arrows indicate approximate molecular weights in thousands of daltons at various points on the gel.



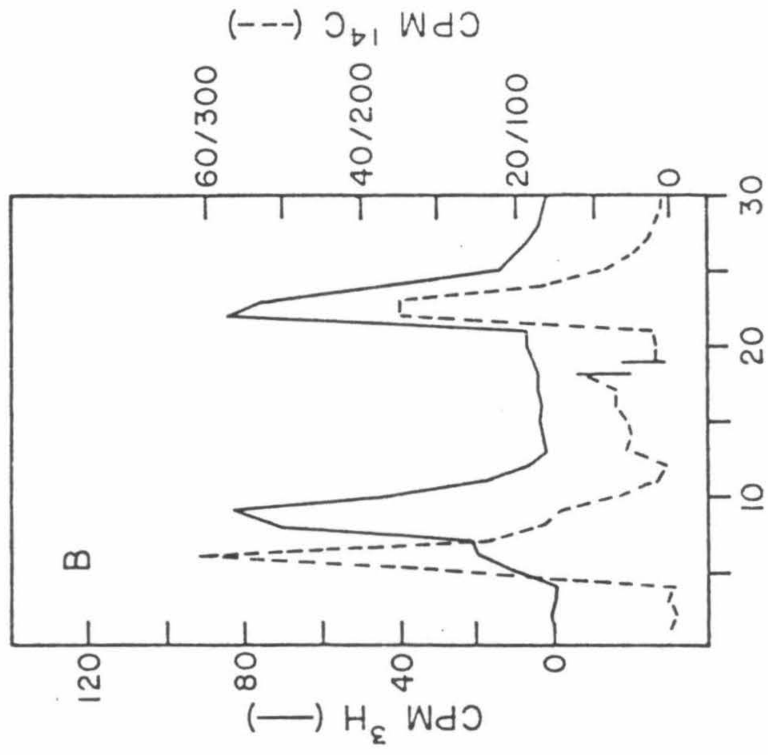
11K
30K
3.1K
CPM ^{3}H (—)
90
60
30
0
CPM ^{14}C (---)
300
200
100
0
20
40
60
FRACTION NUMBER

digest on a strong cation exchanger (Dowex 50W, H^+ form). Under the chromatographic conditions used, a blocked peptide with no lysine or arginine, i.e., a peptide with no cationic groups, should not be bound to the resin and should be found in the column wash through fraction (Narita et al., 1975). Fig. 3a shows the chromatogram obtained. The sample applied to the column contained ^{14}C-methionine labeled capsid protein in addition to the ^{3}H-acetate labeled protein, and the ^{14}C scale has been expanded for the first half of the chromatogram. Thirty percent of the ^{3}H-acetate derived label is found in the column flow through, while a small but significant amount (6%) of the methionine label is also unbound.

We next analyzed the peptides contained in the flow through fraction of the Dowex 50W column by silica gel chromatography and reversed phase HPLC. A silica gel chromatogram of the Dowex 50W flow through of Fig. 3a is shown in Fig. 4a, and as can be seen, the ^{3}H-acetate and ^{14}C-methionine labels comigrate as three peaks. Peak III was identified by co-chromatography with synthetic markers as N-acetyl methionine (results not shown), and hence it must be the modified N-terminal amino acid of the protein. Peak I was found in varying amounts, usually much smaller than in the chromatogram shown in Fig. 4a. In addition, while peak II migrates between methionine and N-acetyl methionine, in other experiments, peak I was found to migrate between the oxidized (sulfoxide) forms of these two species. We therefore believe that peak I is an oxidation product of peak II, containing a methionine sulfoxide residue.

An HPLC chromatogram of these acetate-labeled peptides is shown in Fig. 4b. In this experiment, ^{3}H-acetate labeled capsid protein which became oxidized during purification was digested with chymotrypsin, chromatographed on Dowex 50W and the flow through mixed with ^{35}S-labeled synthetic markers and subjected to HPLC chromatography. The major tritium labeled peak comigrates with N-acetyl methionine sulfoxide, while a minor peak of tritiated N-acetyl methionine is also

FIG. 3. Dowex 50W chromatography of chymotryptic digests of chick embryo fibroblast grown Sindbis proteins. The large peaks at fraction 22 are the peptides eluted from the resin with NaOH. Two scales are given for the ^{14}C label. The left (more sensitive) scale applies to the first 18 fractions, while the right (less sensitive) scale applies to fractions 19-30. (a) Peptides derived from tritiated acetate and ^{14}C-methionine labeled capsid protein. (b) Peptides derived from tritiated acetate and ^{14}C-alanine labeled PE2.


B
CPM ³H (—)
CPM ¹⁴C (---)
120
80
40
0
60/300
40/200
20/100
0
10
20
30

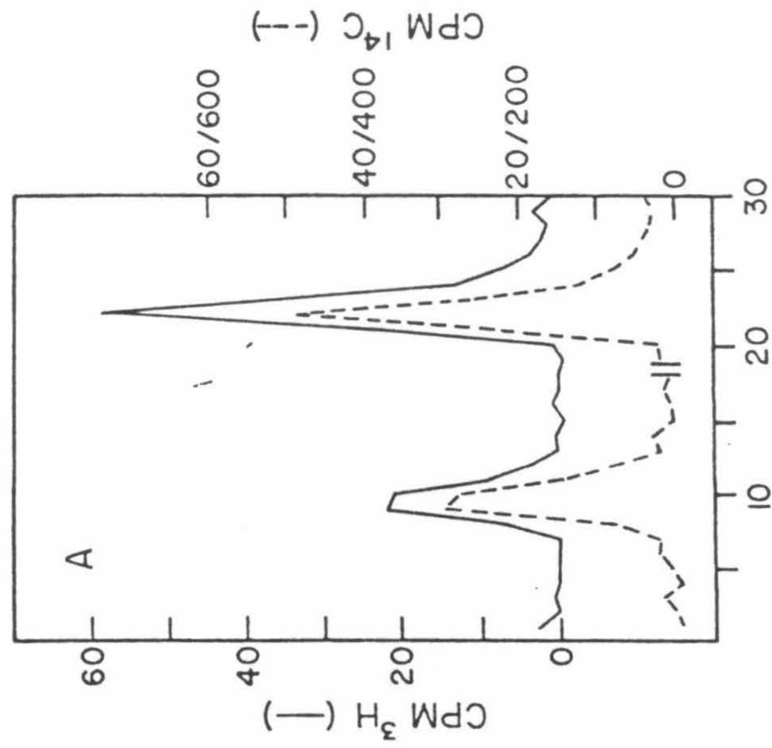
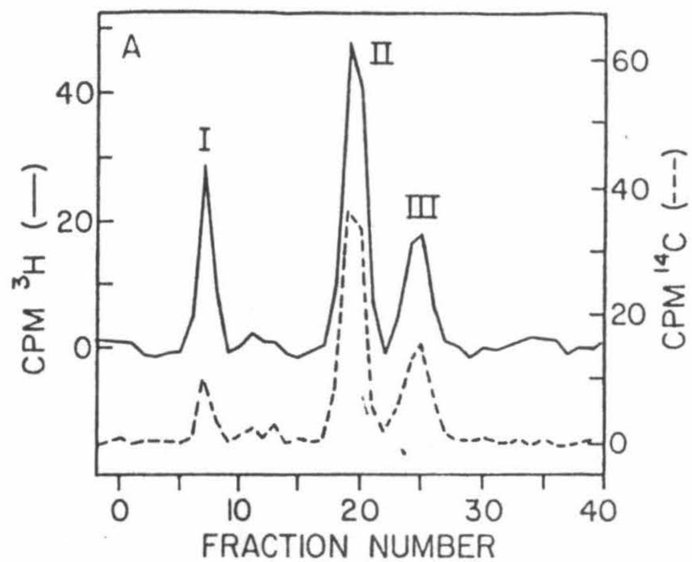
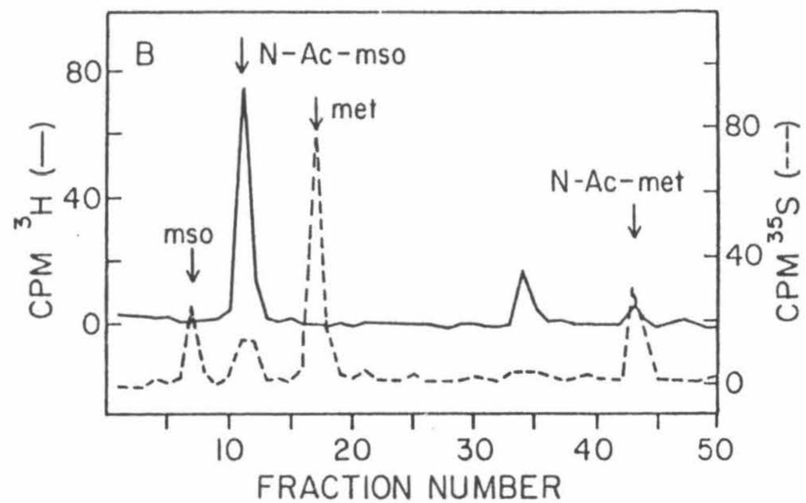
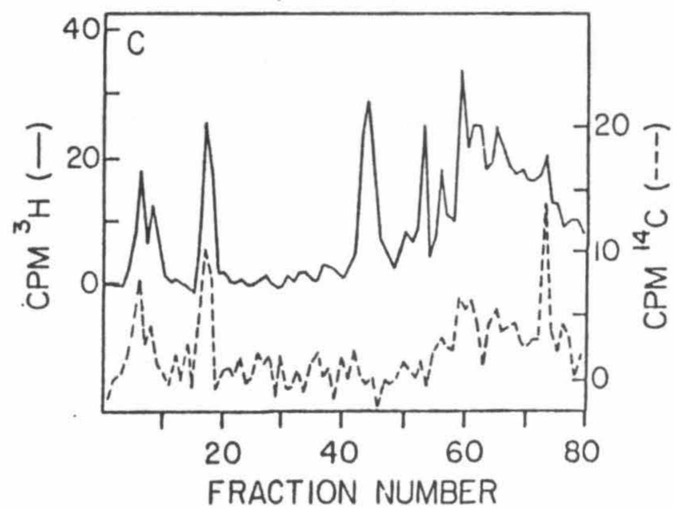
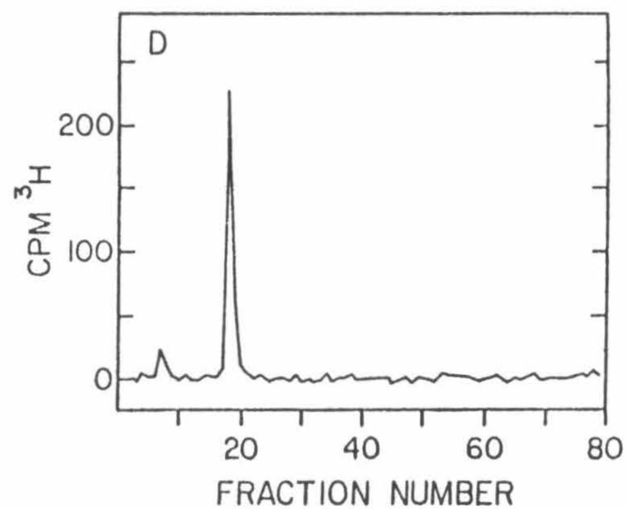

A
CPM ³H (—)
CPM ¹⁴C (---)
60
40
20
0
60/600
40/400
20/200
0
10
20
30

FIG. 4. Analysis of peptides unbound during Dowex 50W chromatography. (a) Silica gel chromatography of tritiated acetate and ^{14}C-methionine labeled peptides derived from the capsid protein. The origin is located at fraction 0, and migration is from left to right, with the solvent front at fraction 38. (b) Reversed phase HPLC of tritiated acetate labeled peptides derived from the capsid protein and ^{35}S labeled markers indicated in the figure as mso (methionine sulfoxide), N-Ac-mso (N-acetyl methionine sulfoxide), met (methionine), and N-Ac-met (N-acetyl methionine). (c) and (d) Reversed phase HPLC of peptides derived from PE2 and labeled with: (c) ^{3}H-proline and ^{14}C-alanine (d) tritiated acetate.


A
I
II
III
CPM ^{3}H (—)
CPM ^{14}C (---)
FRACTION NUMBER


B
N-Ac-mso
met
mso
N-Ac-met
CPM ^{3}H (—)
CPM ^{35}S (---)
FRACTION NUMBER


C
CPM ^{3}H (—)
CPM ^{14}C (---)
FRACTION NUMBER


D
CPM ^{3}H
FRACTION NUMBER

seen. The peak at fraction 34 in the HPLC chromatogram, Fig. 4b, is probably the peak labeled II in Fig. 4a.

In order to further characterize the peptide called peak II above, it was purified from chymotrypsin digests of chemically pure (as opposed to radiochemically pure) capsid protein by HPLC of the Dowex 50W unbound fraction. HCl hydrolysis followed by amino acid analysis established that this peptide contained only aspartic acid or asparagine in addition to the methionine detected by radiolabeling techniques. (Tryptophan would not have been detected.) Although we have not directly determined the structure of this peptide, it almost certainly is N-acetyl-met-asx, also derived from the N-terminus of the protein. It contains the same amount of tritium label per methionine residue as the N-acetyl methionine, and it was isolated by Dowex 50W chromatography, selective for blocked peptides. The presence of asparagine rather than aspartic acid is suggested by the observation that this peptide is not generated by cleavage of the capsid protein with S. aureus V8 protease under conditions in which the peptide bonds C-terminal to aspartic acid and glutamic acid are cleaved (see Fig. 2, and unpublished data).

Assuming that peak II is indeed N-acetyl-met-asn derived from the N-terminal end of the capsid protein, we can estimate the fraction of the tritium label in the N-terminal acetyl group of the capsid protein from the fractions of the input tritium and ^{14}C labels recovered in the Dowex 50W unbound fraction. Since 6% of the applied ^{14}C-methionine is found in the Dowex 50W unbound fraction, and there are 10 methionines in the capsid protein (Bell et al., 1979), the yield of acetylated N-terminal methionine in Fig. 3a is 60%. The recovery of 30% of the tritium label in this fraction implies that about 50% of the tritium label is in this group in the intact protein. Comparable results (40% of the tritium label in the N-terminal N-acetyl methionine) are obtained from a similar analysis of the amounts of the two labels in the N-acetyl methionine peak of the silica gel chromatogram of Fig. 4a.

Acetate-labeled peptides of PE2

The observation that PE2 is labeled with acetate to the same extent as the capsid protein, and thus appears to be acetylated also, led us to examine the Dowex 50W unbound fraction of a proteolytic digest of this protein. Fig. 3b shows the results of Dowex 50W chromatography of the chymotrypsin digestion products of tritiated acetate and ^{14}C-alanine labeled PE2. Two unbound peaks are partially resolved. A large fraction of the tritium label and about 5% of the ^{14}C-alanine label is found in a flow through peak centered at fraction nine, which appears to be the included volume of the column and is the same position as the flow through peak for the digested capsid protein in Fig. 3a. This material was pooled and analyzed by HPLC and silica gel chromatography (see below). The large alanine-containing peak centered at fraction six probably contains peptides which are either large or exist in solution as aggregates so that they cannot enter the pores of the resin, as this is the expected position of the excluded volume of the column. We have also found that the peptides in this early peak do not move from the origin in silica gel chromatography (data not shown).

Figure 4c and d, shows the HPLC chromatogram of the pooled Dowex 50W included volume flow through peaks of the chymotrypsin digestion products of PE2 radiolabeled with ^{3}H-proline and ^{14}C-alanine (Fig. 4c) or tritiated acetate (Fig. 4d). One major acetate labeled peptide is present, and it is also labeled by ^{3}H-proline and ^{14}C-alanine. Many of the additional peptides which elute only with a high organic solvent concentration in Fig. 4c are probably large peptides contaminating the included volume pool from the Dowex 50W chromatography (see Fig. 3b). Using silica gel chromatography instead of HPLC, we also find that an alanine-labeled peak comigrates with the acetate-labeled peak when the two labeled preparations are mixed before chromatography (data not shown).

We next purified 80 pmole of the major acetate-labeled peptide by reverse

phase HPLC from the Dowex 50W unbound fraction of a chymotryptic digest of chemically pure PE2. The peptide was found to have the amino acid composition ($ala_{1.9}$, $leu_{1.0}$, $pro_{1.0}$, $ser_{0.8}$), without correcting for the decomposition of serine during HCl hydrolysis. Since there are about 40 alanine residues in PE2 (J. R. Bell, unpublished data), if we assume this peptide contains two alanines, it is obtained from chymotrypsin digestion and Dowex 50W chromatography in nearly 100% yield, and it contains 45% of the total tritium label in PE2. We also found that this peptide is not labeled if derived from PE2 labeled in vivo with tritiated palmitate (which labels the covalently coupled lipid of the protein) or tritiated glucosamine (results not shown).

We therefore conclude that this PE2-derived peptide is N-acetylated, and thus derived from the N-terminus of PE2. This in turn implies that PE2, or at least a significant fraction of this protein, is blocked in the infected cell.

DISCUSSION

Before analyzing the incorporation of tritiated acetate into the structural proteins of Sindbis virus, we considered it desirable to suppress the metabolic conversion of exogenously supplied acetate into other chemical forms. We estimate the specificity of the acetate labeling, the fraction of the label not metabolically converted to other chemical forms in the viral proteins, to be 40-50% in chick cells labeled with the procedure which we developed. This is based on the amount of label recovered in two overlapping N-acetylated peptides derived from the capsid protein, and the amount of label recovered in an N-acetylated peptide derived from PE2. We also recovered 70% of the acetate derived label in chick grown capsid protein in one 11,000-dalton fragment, which must therefore contain the acetylated (N-terminal) residues of this protein. The additional tritium label in the large fragment is due to the fact that it contains 70% of the proline and 50% of the glutamic acid plus glutamine of the intact protein (unpublished observations). We note that the

specificity of the label is such that it can be used to identify large proteolytic fragments containing the N-terminus of the capsid protein as well as small fragments of one or two amino acids. We have not directly examined the distribution of label in BHK-21 grown, acetate labeled Sindbis proteins. However, the extent of the difference in the labeling of E1, compared with PE2 and the capsid protein, suggests that the specificity of the acetate label may be even greater in this system.

Chromatography on Dowex 50W-X2 of proteolytic digestion products was first used to purify the N-terminal peptide of a blocked protein in studies on the tobacco mosaic virus (TMV) coat protein. Narita (1958), in his classic studies, reasoned that a blocked peptide, if it contained no lysine or arginine, could not be bound to a strong cation exchanger, while all other peptides in the digest should contain at least the N-terminal amino group and so should have an affinity for the resin. Since that time, a number of authors have used this technique (Ward and Dopheide, 1980; see also Narita et al., 1975). The use of radiolabel, in particular radioactive acetate in the case of an acetylated N-terminus, has allowed us to extend this technique to situations in which it is difficult to obtain sufficient material for the usual biochemical analyses. In the case of Sindbis virus infected cells, it is easy to obtain radiochemically pure viral proteins and their precursors since the virus shuts down host cell protein synthesis (see, for example, Fig. 1). It has also been found that most of the contaminating peptides which are present in the Dowex 50W-X2 unbound fraction are relatively large, as they are also not retained by the anion exchanger Dowex 1-X2 (Narita and Ishii, 1962). The use of radiolabeled acetate has allowed us to extend the discrimination of the Dowex 50W chromatographic separation, for we can distinguish and partially separate those small peptides which enter the pores of the resin and are not bound, from peptides which are not retained on the column simply because they are excluded from the pores of the ion exchanger.

This determination of the N-terminal sequence of the capsid protein as N-acetyl-met-asn- is consistent with the results of other workers. Methionine is the N-terminal amino acid of the capsid protein when it is synthesized in vitro in the presence of an enzymatic system which removes acetyl-CoA from the reaction mixture (Bonatti and Blobel, 1979), and a blocked peptide with the amino acid composition (asx, met, arg) has been isolated from tryptic digests of the capsid protein (Boege et al., 1980). Finally, our deduced sequence is consistent with the sequence of the viral mRNA which codes for the capsid protein (C. M. Rice and J. H. Strauss, manuscript in preparation).

Experiments on the incorporation of acetate label into PE2 led us to conclude that a peptide derived from this protein, with the amino acid composition ($ala_{1.9}$, $leu_{1.0}$, $pro_{1.0}$, $ser_{0.8}$), is N-acetylated and thus derived from the amino-terminus of N-acetylated PE2. This result is surprising in view of the fact that a partial N-terminal amino acid sequence has been obtained by Edman degradation of radiolabeled PE2 (Bonatti and Blobel, 1979). We have obtained a complete N-terminal sequence (ser-ala-ala-pro-leu-) by Edman degradation of PE2 which is consistent with the previously determined partial sequence, but the recovery of this sequence indicated that 80-90% of the protein was blocked, and it was shown to be unlikely that the blocking occurred during purification (Chapter 5, this thesis). The composition of the N-acetylated peptide which we have isolated is consistent with the proposal that it was obtained from a blocked fraction of PE2 which is related to the sequenceable fraction by N-acetylation. We also note that by far the most common acetylated amino acid in N-acetylated proteins is serine (Jörnvall, 1975; Narita et al., 1975). We describe the N-terminus of PE2 in more detail in a separate report (Chapter 5, this thesis).

One feature of the list of proteins which are known to be blocked (Narita et al., 1975; Jörnvall, 1975) is that these proteins are often ones which the cell would not be

expected to turn over and resynthesize extensively. Some examples are actin and cytochrome c, both from a variety of sources, and keratin from sheep wool and emu feather. The nature of these blocked proteins suggests that the lack of a free N^{α}-amino group may have some role in the protection of these proteins from proteolytic digestion, presumably initiated by exopeptidases (Jörnvall, 1975; Brown and Roberts, 1976). One might then expect viral proteins to be often blocked, since it is obviously advantageous to a virus that its structural proteins not be degraded. In fact, blocked structural proteins have been found in TMV and other plant viruses (Narita et al., 1975; Jörnvall, 1975), Mengo virus (Ziola and Scraba, 1976), Sendai virus (Scheid and Choppin, 1977), murine leukemia virus (Oroszlan et al., 1978), and influenza virus (Ward and Dopheide, 1980). In the case of adenovirus, of the five structural proteins of known N-terminal sequence, four are blocked (Jörnvall et al., 1974; Bowdin et al., 1979; Anderson and Lewis, 1980), and the other is produced by proteolytic cleavage from a blocked precursor (Sung et al., 1977). Thus the determination of the N-termini of acetylated proteins may be a problem encountered particularly often in the sequencing of viral proteins, and in vivo labeling of such proteins with radioactive acetate may frequently be useful in the elucidation of their primary structures.

ACKNOWLEDGEMENTS

We are grateful to Dr. L. E. Hood and his colleagues, in particular Dr. M. W. Hunkapiller, for the use of their equipment and expertise in the separation of peptides by HPLC and the determination of amino acid compositons; to C. M. Rice for his scintillating discussions during the course of this work; and to E. M. Lenches for expert technical assistance and conscientiously growing large amounts of unlabeled virus. This work was supported by grants GM06965 and AI10793 from the National Institute of Health and by grant PCM77-26728 from the National Science Foundation. J.R.B. was supported by training grant GM00086 from the National Institute of Health.

REFERENCES

Anderson, C. W., and Lewis, J. B. (1980). Amino-terminal sequence of adenovirus type 2 proteins: hexon, fiber, component IX, and early protein 1B-15K. Virology **104**, 27-41.

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. Nat. 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.

Boege, U., Wengler, G., Wengler, G., and Wittmann-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.

Boudin, M.-L., Moncany, M., D'Halluin, J.-C., and Bowlanger, P. A. (1979). Isolation and characterization of adenovirus type 2 vertex capsomer (penton base). Virology **92**, 125-138.

Brown, J. L., and Roberts, W. K. (1976). Evidence that approximately 80% of the soluble proteins from Ehrlich ascites cells are N^{α}-acetylated. J. Biol. Chem. **251**, 1009-1014.

Drapeau, G. R., Boily, Y., and Houmard, J. (1972). Purification and properties of an extracellular protease of Staphylococcus aureus. J. Biol. Chem. **247**, 6270-6726.

Eagle, H. (1959). Amino acid metabolism in mammalian cell cultures. Science **130**, 432-437.

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.

Ikenaka, T., Bammerlin, H., Kaufmann, H., and Schmid, K. (1966). The amino-terminal peptide of α_1-acid glycoprotein. J. Biol. Chem. **241**, 5560-5563.

Jörnvall, H. (1975). Acetylation of protein N-terminal amino groups. Structural observations on α-amino acetylated proteins. J. Theor. Biol. **55**, 1-12.

Jörnvall, H., Ohlsson, H., and Philipson, L. (1974). An acetylated N-terminus of adenovirus type 2 hexon protein. Biochem. Biophys. Res. Commun. **56**, 304-310.

Jörnvall, H., Von Bahr-Lindstrom, H., and Philipson, L. (1978). Radioactive labeling of acidic regions in the adenovirus hexon protein through metabolic conversion of [^{14}C] acetate. FEBS Lett. **88**, 237-241.

Kolb, J. J., and Toennies, G. (1942). The investigation of amino acid reactions by methods of non-aqueous titrimetry I. Acetylation and formylation of amino groups. J. Biol. Chem. **144**, 193-201.

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

Narita, K. (1958). Isolation of acetylpeptide from enzymatic digests of TMV-protein. Biochem. Biophys. Acta **28**, 184-191.

Narita, K., and Ishii, J. (1962). N-terminal sequence in ovalbumin. J. Biochem. (Tokyo) **52**, 367-373.

Narita, K., Matsuo, H., and Nakajima, T. (1975). End group determination. In "Protein Sequence Determination", 2nd edition (S. B. Needleman, ed.), pp. 30-103. Springer-Verlag, New York.

Oroszlan, S., Henderson, L. E., Stephenson, J. R., Copeland, T. D., Long, C. W., Ihle, J. N., and Gilden, R. V. (1978). Amino- and carboxyl-terminal amino acid sequences of proteins coded by gag gene of murine leukemia virus. Proc. Nat. Acad. Sci. USA **75**, 1404-1408.

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.

Scheid, A., and Choppin, P. W. (1977). Two disulfide-linked polypeptide chains constitute the active F protein of paramyxoviruses. Virology **80**, 54-66.

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

Sefton, B. M., and Keegstra, K. (1974). Glycoproteins of Sindbis virus: preliminary characterization of the oligosaccharides. J. Virol. **14**, 522-530.

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

Sung, M. T., Lischove, M. A., Richards, J. C., and Hosokawa, K. (1977). Adenovirus chromatin I. Isolation and characterization of the major core proteins VII and precursor pro-VII. J. Biol. Chem. **252**, 4981-4987.

Swank, R. T., and Munkres, K. D. (1971). Molecular weight analysis of oligopeptides by electrophoresis in polyacrylamide gel with sodium dodecyl sulfate. Analyt. Biochem. **39**, 462-477.

Ward,C. W., and Dopheide, T. A. A. (1980). Completion of the amino acid sequence of a Hong Kong influenza hemagglutinin heavy chain: sequence of cyanogen bromide fragment CN1. Virology **103**, 37-53.

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.

Ziola,B. R., and Scraba, D. G. (1976). Structure of the Mengo virion IV. Amino- and carboxy-terminal analysis of the major capsid polypeptides. Virology **71**, 111-121.

Chapter 5

The N-Terminus of PE2 in Sindbis Virus Infected Cells

JOHN R. BELL, CHARLES M. RICE, MICHAEL W. HUNKAPILLER, and
JAMES H. STRAUSS

This chapter has been prepared for submission to the
Journal of Molecular Biology

ABSTRACT

Antibodies specific for each of the structural proteins of Sindbis virus were prepared and characterized, and antibody directed against E2, one of the two envelope glycoproteins, was used to purify microgram quantities of intracellular E2 and its precursor, PE2, from infected cells. PE2 and E2 were found to constitute a significant fraction of the total protein in infected cells at an intermediate time after infection. Two N-termini of PE2 were found, and one, the primary product of translation, was sequenced by Edman degradation. The second N-terminus, found in 80-90% of PE2, is blocked and is produced by N-acetylation of the first. The N-terminus of PE2, which is probably the signal sequence of this protein, also contains a glycosylated asparagine residue. The putative signal sequence is not cleaved by signal peptidase, either under normal conditions or when glycosylation is prevented by tunicamycin.

INTRODUCTION

Two of the three structural proteins of Sindbis virus, E1 and E2, are membrane glycoproteins, each with a molecular weight about 50,000 daltons (Strauss and Strauss, 1977). In the virion, most of the mass of these two proteins is external to the lipid bilayer of the virus envelope (Compans, 1971), although each is anchored in the bilayer by short stretches of hydrophobic amino acids near the C-terminal ends of the proteins (Garoff and Soderlund, 1978). The structure of these proteins has also been examined after translation in an *in vitro* system containing microsomes. In this system, both E1 and PE2, the precursor to E2, are inserted into membrane vesicles and this translocation is cotranslational (Garoff *et al.*, 1978; Bonatti *et al.*, 1979). The third viral structural protein, the capsid protein (mw 30,000 daltons), forms a complex with the RNA of the virus. This complex, the viral capsid, is found internal to the lipid envelope in the virion (Compans, 1971), and in the cytosol of infected cells (Wirth *et al.*, 1977).

All three structural proteins are translated from a single mRNA, which contains only one initiation site (Cancedda *et al.*, 1975), and translation results in a polyprotein product which is processed by proteolytic cleavage while still nascent (Strauss and Strauss, 1977; Garoff *et al.*, 1978). This mRNA contains, in order from the 5' end, the sequences for the capsid protein, PE2, and E1 (Welch and Sefton, 1979). The capsid protein is first cleaved off the growing polypeptide chain, as soon as its translation is complete, and the nascent PE2 is immediately inserted into the lumen of the endoplasmic reticulum (Garoff *et al.*, 1978; Bonatti *et al.*, 1979). The cleavages between the sequences of PE2 and E1, and the insertion of E1 into the membrane also occur before the translation of E1 is completed (Garoff *et al.*, 1978), but are less well understood. As a final step in the maturation of the virus, about 20 min after its synthesis PE2 is cleaved to yield E2, the mature component of the Sindbis virion (Schlesinger and Schlesinger, 1972).

We are interested in the primary sequence of the viral structural proteins (Bell *et al*., 1978; Rice and Strauss, manuscript in preparation), but the N-terminus of PE2 is of particular interest, for the insertion of the major portion of PE2 into the lumen of the endoplasmic reticulum has been shown to be mediated by the N-terminal portion of the protein (Garoff *et al*., 1978). Proteins secreted from eukaryotic cells, with one exception, have been found to contain an N-terminal extension of 15-30 amino acids which, for a short time after its synthesis and while most of the rest of the protein is still untranslated, is able to interact with the rough endoplasmic reticulum. The interaction of this extension, or signal sequence, with the membrane results in the translocation of the protein across the membrane and into the lumen of the rough endoplasmic reticulum, and this translocation is a prerequisite to further processing of the protein (for review, see Blobel *et al*., 1979). This model has been extended (Blobel and Doberstein, 1975) to describe the translocation of the N-terminal portions of some integral membrane proteins via a similar N-terminal signal sequence, also cleaved after translocation, and the G protein of vesicular stomatitus virus has been shown to conform to the model (Lingappa *et al*., 1978). The same method of insertion of membrane proteins has also been found in bacteria, although several exceptions are known both in prokaryotes and eukaryotes (see Davis and Tai, 1980; Blobel *et al*., 1979). The N-terminal region of PE2 of Sindbis virus (or the closely related Semliki Forest virus) also functions as a signal sequence in translocation, for in an *in vitro* translation-translocation system PE2 is inserted into membrane vesicles only if the membranes are added before the first 50 amino acids of the protein have been synthesized (Garoff *et al*., 1978). The observation that the C-terminal 30-40 amino acids of a nascent peptide are sequestered by the ribosome (Malkin and Rich, 1967; Blobel and Sabatini, 1970) further localizes this translocation function to the extreme N-terminus.

A partial N-terminal sequence has been obtained from radiolabeled PE2 in

which 11 of the first 25 amino acids were identified, and this partial sequence was not inconsistent with its role as a signal sequence for the translocation of the protein across the membrane of the endoplasmic reticulum (Bonatti and Blobel, 1979). We now report the preparation of chemically pure PE2 (and also its product E2) from Sindbis infected cells and the extensive characterization of its N-terminus. During the course of this work, we found heretofore unknown features of this important region of the protein which distinguish it from other known signal sequences.

MATERIALS AND METHODS

Virus strain

The virus used in these studies was the large plaque variant of the HR strain of Sindbis virus, originally obtained from Dr. B. W. Burge and plaque purified several times in our laboratory.

Preparation of antisera

Sindbis virus was purified and structural proteins were prepared as a solution in sodium dodecyl sulfate (SDS) as previously described (Bell *et al.*, 1979). The purified proteins were precipitated by the addition of two volumes of 100% ethanol, frozen on dry ice or incubated overnight at -20°C, warmed to 4°C, and centrifuged at 15,000 rpm for 15 min at 4°C in a Sorvall SS34 rotor. The supernatants were discarded and the pellets air-dried and resuspended in 10 m*M* Tris-Cl, 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 an 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. A crude gamma globulin fraction was prepared by ammonium sulfate fractionation as described by Garvey et al. (1977). For the purification of IgG, this crude gamma globulin was first dialyzed against 10 mM $NaPO_4$, pH 7.2, 15 mM NaCl. After centrifugation at 10,000 x g, for 20 min at 4°C, the supernatant was passed over a column of CM52, then DE52 (Whatman) equilibrated in 10 mM $NaPO_4$ pH 7.2, 15 mM NaCl (Palacios et al., 1972). The excluded IgG containing fractions were collected, pooled, and stored frozen at -70°C.

Preparation of radiolabeled infected cell lysates and virus

Primary chick embryo fibroblasts were prepared and infected with Sindbis virus at a multiplicity of infection of 20-40 as described (Pierce et al., 1974). Purified virus was obtained from infected cell monolayers labeled overnight with ^{35}S-methionine and harvested in high salt (Pierce et al., 1974), and purified by rate zonal centrifugation followed by isopycnic sedimentation according to Bell et al. (1979). Infected cell monolayers were labeled and harvested in 0.5-1.0% SDS as described elsewhere (Chapter 4, this thesis), but this procedure was modified for specific purposes as follows. For pulse-labeling with ^{35}S-methionine, the cells were washed twice with methionine-free medium just before the addition of the labeling medium, which contained no nonradioactive methionine, and in some cases the labeled proteins were chased by the replacement of the medium with medium containing no radioactivity but unlabeled methionine at twice the normal concentration, followed by an additional 20 min incubation before harvest. Monolayers treated with tunicamycin (the generous gift of Dr. Robert L. Hamill, Lilly Research Laboratories) were incubated in medium containing 0.5 μg/ml of the drug for 4.5 to 6.5 hr post-infection, and labeled from 6.5 to 7.5 hr post-infection in the absence of the drug. The procedure for in vivo incorporation of ^{3}H-acetate is described in detail elsewhere (Chapter 4, this thesis).

Immune-precipitation

In preparation for immune-precipitation, radiolabeled infected cell monolayers (dissolved in 0.5% SDS) were heated to 56°C for 10 min and diluted and adjusted to 0.05 M Tris-Cl, pH 7.4, 0.2 M NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.5% Triton X-100, and 1 mg/ml bovine serume albumin (TNA buffer). After dilution, there was at least a five-fold excess of Triton X-100 over SDS by weight, and immune-precipitation was performed directly after Triton X-100 addition, without freezing. Radiolabeled purified virus samples were similarly denatured by the addition of SDS to a concentration of 0.5% and heating as above, and diluted into TNA. Alternatively, the heating step was omitted and an equivalent amount of SDS was added after the addition of Triton X-100 or omitted entirely.

Rabbit IgG was diluted into TNA before being used at a final concentration of about 100 μg/ml. Incubation of antigen and antibody was for 30 min at room temperature, 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. Immune-precipitates were washed with 0.05 M Tris-Cl, pH 7.4, 0.2 M NaCl, 1 mM EDTA, 0.1% Triton X-100, 200 μg/ml bovine serum albumin, and dissolved in sample buffer and analyzed by SDS-polyacrylamide gel electrophoresis. Before loading on the gel, samples were heated as above and clarified by centrifugation at 3500 x g for 6 min at room temperature.

SDS-polyacrylamide gel electrophoresis

The buffer system of Laemmli (1970), except containing half the concentration of Tris in the gels, was used. For analytical purposes, samples were electrophoresed in 6 mm diameter cylindrical gels, which were then stained with Coomassie blue or sliced into 1 mm fractions and counted by liquid scintillation. Alternatively, samples

were electrophoresed on slab gels, which were dried and analyzed by fluorography according to Bonner and Laskey (1974). Samples for preparative gel electrophoresis were precipitated with ethanol as above and electrophoresed in 8 mm diameter cylindrical gels containing 0.01% mercaptopropionic acid in the sample and upper electrode buffer. After electrophoresis, the gel was sliced into approximately 100 fractions, each consisting of two 0.5 mm slices, and each fraction was eluted in 1 ml of 0.5% SDS, 10 mM 2-(N-morpholino)ethanesulfonic acid (MES), 2.9 mM NaOH, 1 mM dithiothreitol, 20 µg/ml phenylmethylsulfonyl fluoride. The pH of this buffer, after equilibration with the gel slices, was approximately 6.4. After shaking for 9 hr at room temperature, an aliquot of each fraction was assayed for radioactivity by liquid scintillation counting. Appropriate fractions were pooled, clarified by centrifugation at 2000 x g, and the supernatant volume reduced 10-fold by lyophilization.

Immuno-affinity chromatography

Ten roller bottles of primary chick embryo fibroblasts (800 cm^2 surface area and 10^8 cells per roller bottle) were infected as described above, and one was labeled by the addition of ^{14}C-alanine to the medium 6-7 hr post-infection. Seven hours post-infection, the monolayers were chilled on ice, washed twice with ice-cold phosphate buffered saline (PBS), and the cells were scraped from the glass with a rubber policeman in a small volume of PBS, pooled, pelleted by centrifugation at 500 x g for 3 min at 4°C, and washed once more with cold PBS. After resuspension in 15 ml of ice-cold 0.05 M Tris-Cl, pH 7.5, 0.2 M NaCl (TN buffer) plus 20 µg/ml phenylmethylsulfonyl fluoride, the cells were lysed by the addition of SDS to 2%, heated to 70°C for 20 min, and centrifuged at 100,000 x g for 1 hr at room temperature to pellet DNA. The supernatant was precipitated with ethanol as above, and precipitated protein was redissolved in 7.5 ml of 1% SDS, 0.05 M Tris-Cl, pH 7.5, and diluted and adjusted on ice to 2% Triton X-100, 0.05 M Tris-Cl, pH 7.5, and 0.2 M NaCl, to a total volume of 35 ml, and applied to the immunoabsorbant.

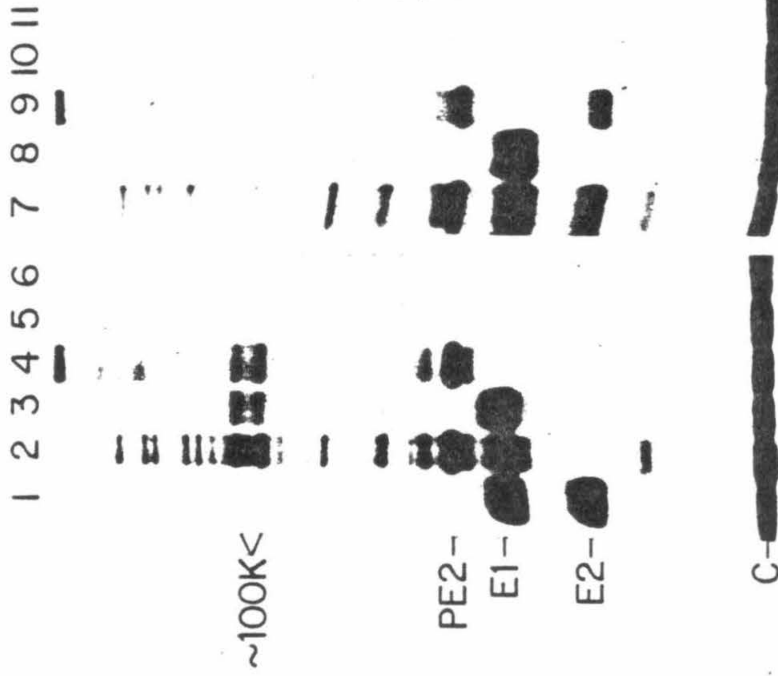
Solid phase antibodies were prepared from the crude gamma globulin fraction of anti-E2 serum, which was first dialyzed against 10 m<u>M</u> $NaPO_4$, pH 7.5, and clarified by centrifugation at 10,000 x g for 20 min at 4°C. The supernatant (640 mg of protein in 26 ml) was gently mixed with 28 ml (wet settled volume) of washed, activated crosslinked agarose beads (Reacti-gel 6X, Pierce Chemical Co.) in 0.1 M Na-borate, pH 8.5, for three days at 4°C (Bethell et al., 1979). After washing with 0.1 M Na-borate, pH 8.5, at 40°C, the protein content of the agarose beads was 8.5 mg/ml wet settled volume.

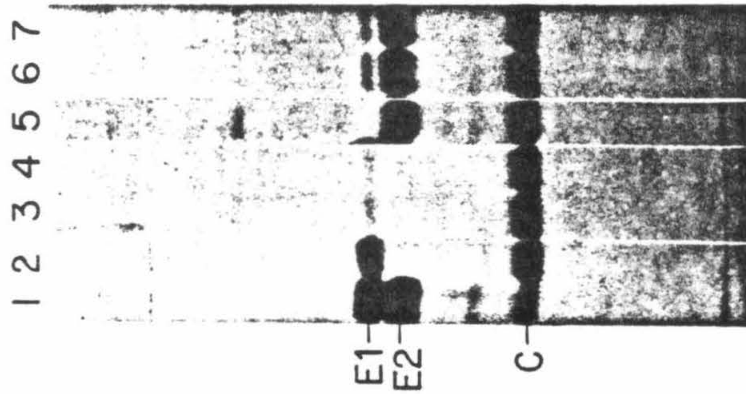
A 1.8 cm diameter column was poured with 28 ml of the immuno-absorbent and run at 4°C at a flow rate of 18 ml/hr. After washing with 10 bed volumes of 1% Triton X-100 in TN buffer, the sample was applied and 10 min fractions were collected. The column was then washed with 1.5 bed volumes of 0.05% Triton X-100 in TN buffer, and warmed and eluted at room temperature with 0.5% SDS, 10 mM Tris, pH 7.5. Aliquots of each fraction were assayed for radioactivity, and fractions containing protein eluted with SDS were pooled (22 ml total), lyophilized, and dissolved in 2.2 ml 0.5% β-mercaptoethanol.

Protein sequenation

Samples were precipitated with ethanol as above, dissolved in a small volume of trifluoroacetic acid, and loaded onto a non-commercial spinning cup sequenator, the construction and operation of which have been described (Hunkapiller and Hood, 1980). When chemically pure protein was sequenced, the mixture of phenylthiohydantoin (PTH) amino acid derivatives released at each cycle was analyzed by reversed phase HPLC (Johnson et al., 1979). For the analysis of radiolabeled samples, tuna heart cytochrome c was added to 20 μg/ml before ethanol precipitation, 1 mg of sperm whale apomyoglobin was added to the spinning cup with the sample, and the mixtures of PTH amino acid derivatives released were dried under nitrogen and counted directly by liquid scintillation without fractionation.

FIG. 1. SDS polyacrylamide gel electrophoresis of the immune precipitate obtained with antibodies to the viral structural proteins. (a) ^{35}S-Methionine-labeled purified virus (lane 1) was precipitated with anti-E1 (lanes 2-4) or anti-E2 (lanes 5-7). Before immune-precipitation, virus samples were denatured with SDS (lanes 2 and 5), or solubilized with Triton X-100 (lanes 3 and 6), or solubilized with Triton X-100 followed by the addition of SDS (lanes 4-7) as described in Materials and Methods. (b) SDS-Denatured infected cell monolayers, pulse-labeled with ^{35}S-methionine for 10 min (lanes 2-6) or chased for 20 min after similar labeling (lanes 7-11), are shown in lanes 2 and 7. Immune-precipitates were obtained from these samples with anti-E1 (lanes 3 and 8), anti-E2 (lanes 4 and 9), anti-capsid (lanes 5 and 10), or non-immune IgG (lanes 6 and 11). Lane 1 contained an untreated virus marker. In the infected cell, E2 is seen only after the chase period, and bands labeled 100 K are discussed in the text.


B
1 2 3 4 5 6 7 8 9 10 11
~100K<
PE2
E1
E2
C


A
1 2 3 4 5 6 7
E1
E2
C

RESULTS

Characterization of antibodies

We first examined the reaction of the antibodies, which had been produced in response to purified, SDS-denatured Sindbis structural proteins, with the individual proteins of the virion. Immune-precipitates were analyzed by SDS polyacrylamide gel electrophoresis, with the results shown in Fig. 1a. As can be seen, the capsid protein was precipitated non-specifically under all conditions examined. When the viral proteins were denatured with SDS and then diluted in an excess of Triton X-100 buffer, anti-E1 (Fig. 1a, lane 2) or anti-E2 (Fig. 1a, lane 5) antibodies precipitated, in addition to the capsid protein, only the corresponding envelope protein. However, when the virus was disrupted not with SDS but with the non-ionic detergent, Triton X-100, which does not denature most proteins (Helenius and Simons, 1975), quite different results were obtained. Under these conditions, anti-E1 antibody was no longer reactive against E1 (Fig. 1a, lane 3), while anti-E2 precipitated a small but significant amount of E1 in addition to E2 (Fig. 1a, lane 6), and in control experiments the same result was obtained when SDS was added after disruption with Triton X-100 (Fig. 1a, lanes 4 and 7). Thus, the different reactivity of the antibodies to the SDS-denatured proteins, when compared with the Triton X-100 solubilized proteins, must be due to changes in the tertiary or quaternary structure of the antigens rather than the physical presence of SDS in the reaction mixture.

We next ascertained the specificities of the antibodies for the viral structural proteins and related proteins in SDS-solubilized infected cell lysates (Fig. 1b). After the addition of anti-capsid antibody, only the capsid protein was precipitated (Fig. 1b, lanes 5 and 10), although this protein again precipitated non-specifically with all antibodies, including non-immune IgG (Fig. 1b, lanes 6 and 11). However, other experiments (using a solid phase immuno-absorbent) have shown that anti-capsid antibody does in fact bind the capsid protein (unpublished observations). The

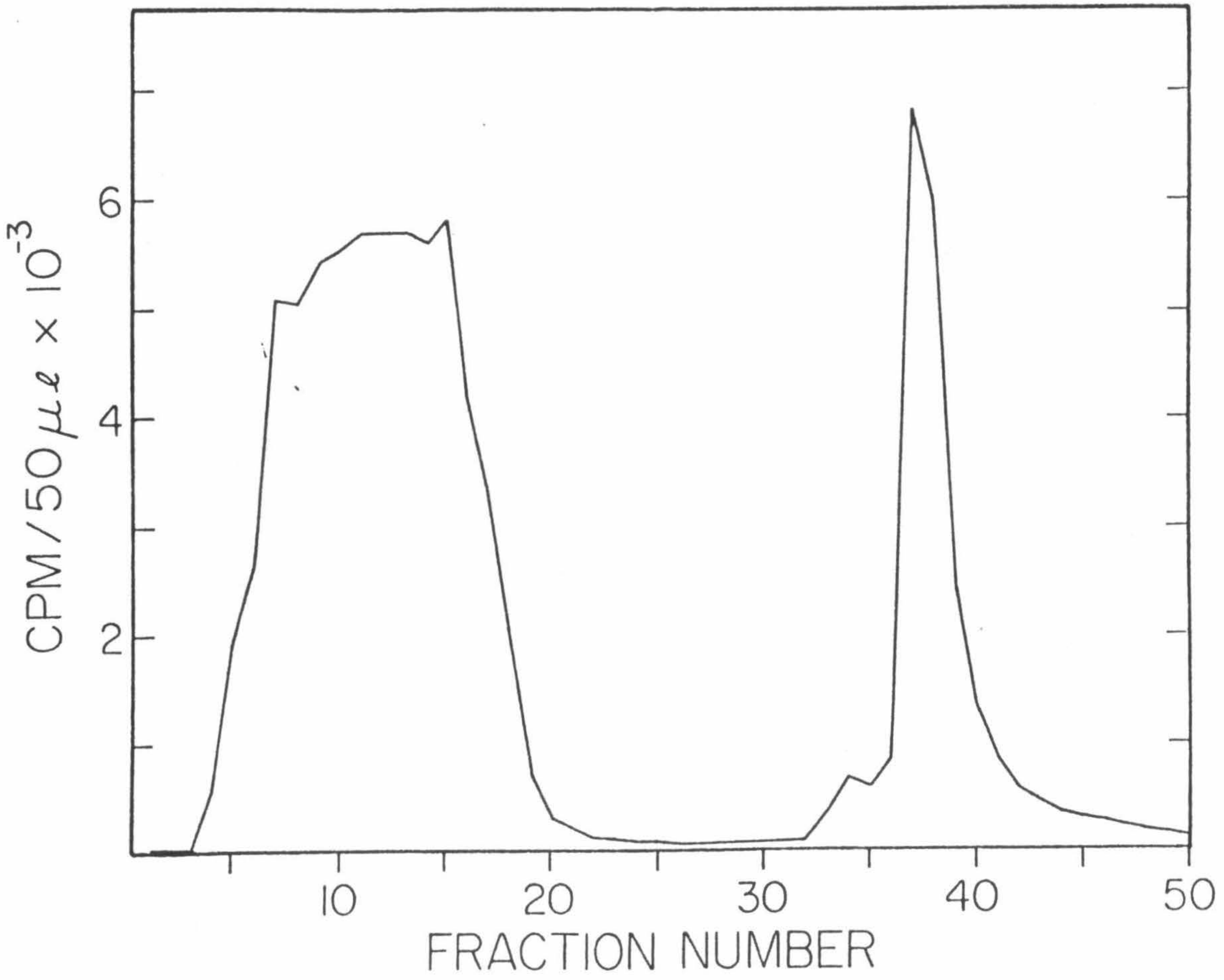
specificities of the antibodies for E1 and E2 are the same as seen when SDS-disrupted virions are used as antigen. In addition, PE2 is precipitated by anti-E2, but not by anti-E1 (Fig. 1b, lanes 4 and 9), and two short-lived proteins, labeled ~100 K proteins after their approximate molecular weights, are specifically precipitated by both anti-E1 and anti-E2 (Fig. 1b, lanes 3 and 4) and so contain antigenic determinants of both envelope proteins. A 100,000 dalton non-glycosylated protein, the B protein, which contains the sequences of both PE2 and E1 and which is probably the result of an error in processing rather than an intermediate in envelope glycoprotein biosynthesis, is found in BHK cells infected with Sindbis virus (see Discussion). However, the relationship of these ~100 K proteins in chick cells to the B protein is unknown.

Purification of PE2 and E2

The availability of antibody specific for PE2 and E2 allowed us to use an immunoaffinity column to purify PE2 from infected cells. Figure 2 shows the pattern obtained when a ^{14}C-alanine labeled infected cell lysate, harvested in SDS and diluted with Triton X-100, was passed over a column containing covalently coupled anti-E2, and Fig. 3. shows the results of SDS polyacrylamide gel electrophoresis used to follow the purification procedure. Less than 5% of the capsid protein was bound to the immunoaffinity column, despite the fact that it precipitated non-specifically during immune precipitation in solution. The specificity of the immune reagent for PE2 and E2, in comparison with host cell proteins, can be inferred from the pattern of the stained gel of the antibody bound fraction, Fig. 3c, as E2 cannot be detected in polyacrylamide gels of infected cell lysates stained for protein.

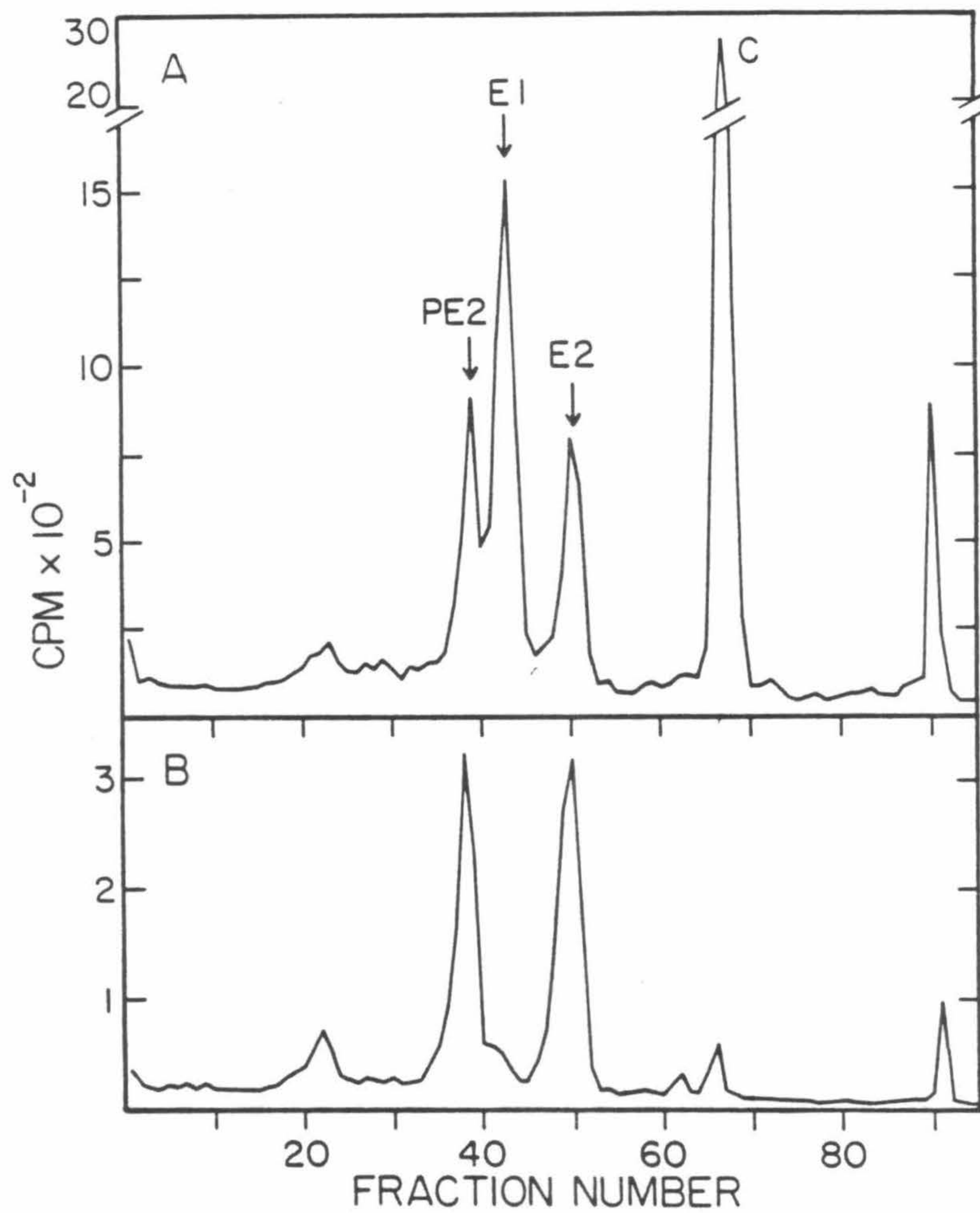
We presume that the level of contaminating protein which is seen in Fig. 3c could have been further reduced by preliminary passage of the cell lysate through a column containing covalently coupled non-immune serum before the immunoaffinity chromatography. However, this was unnecessary since we next separated PE2 from E2, and from the other contaminating proteins, by preparative polyacrylamide gel

FIG. 2. Immuno-affinity chromatography with anti-E2. SDS-Denatured infected cells, labeled with ^{14}C-alanine, were passed over a column containing the covalently coupled antibody, and an aliquot of each fraction was counted for radioactivity, as described in Materials and Methods. The sample was applied in Triton X-100, and the column was washed with a Triton X-100 containing buffer beginning with fraction 12. Elution with SDS was begun at fraction 28, and the fraction size at this point increased from 3.0 to 3.7 ml per fraction.



CPM / 50 μℓ × 10-3
2
4
6
10
20
30
40
50
FRACTION NUMBER

FIG. 3. SDS polyacrylamide gel analysis of the results of immuno-affinity chromatography. The gels in panels a and b were sliced and counted for radioactivity, while the gel in panel c was stained for protein with Coomassie blue. Electrophoresis was from left to right, and the identities of the viral proteins are indicated in panels a and c. Samples consisted of: a) the infected cell lysate, and b) and c) the pooled fractions eluted from the immuno-affinity column with SDS.



C

PE2 E2

electrophoresis. We expected from Fig. 3c, the stained polyacrylamide gel pattern of the antibody-bound fraction, that a preparative application of this technique would result in an essentially pure protein preparation, and this was substantiated by the results of N-terminal amino acid sequencing (see below). To minimize the possibility of modification of N-terminal amino groups, the gel system was modified by the inclusion of a negatively charged mercaptan in the sample and upper reservoir buffer and proteins were eluted from the wet gel immediately after electrophoresis at a slightly acidic pH, at which amino groups are protonated.

After immuno-affinity chromatography and preparative gel electrophoresis, we recovered 130 μg (determined by quantitative amino acid analysis), or 2.2 n moles, of PE2 from 10 roller bottles of infected cells (about 10^9 cells, or 150 mg of protein total). The E2 from the preparative gels was also recovered, and we obtained 270 μg of this protein. The yields of both proteins during the individual steps of the purification, determined by summing the radiolabeled peaks in Figs. 2a and b, were similar. 47% of the PE2 and 52% of the E2 present in the cell lysate was recovered in the fraction eluted from the immunoabsorbant, and overall recoveries were 43% for PE2 and 49% for E2.

N-terminal sequence of PE2

The primary amino acid sequence data obtained by automated Edman degradation of 780 pmoles of the purified PE2 preparation are shown in Fig. 4. The first amino acid in the sequence is probably serine, as extensive destruction of the PTH derivative of this amino acid occurs during sequenation. The rather large background of several other amino acids during the first cycle is consistent with our experience in sequencing such small amounts of protein. The amino acid sequence deduced from the data of Fig. 4 is presented in Fig. 5, which also shows the translated sequence of a portion of the mRNA coding for the Sindbis structural proteins. During Edman degradation, we could not detect the cysteine at position 10, the asparagine at

FIG. 4. Yields of amino acid PTH-derivatives from the spinning cup sequenator analysis of 780 pmoles of PE2. Aliquots of each cycle were analyzed by HPLC, peaks were quantitated by comparison with a standard mixture of PTH-amino acids, and the yields were normalized to an injection of 100% of the sample. PTH-cysteine was not determined.

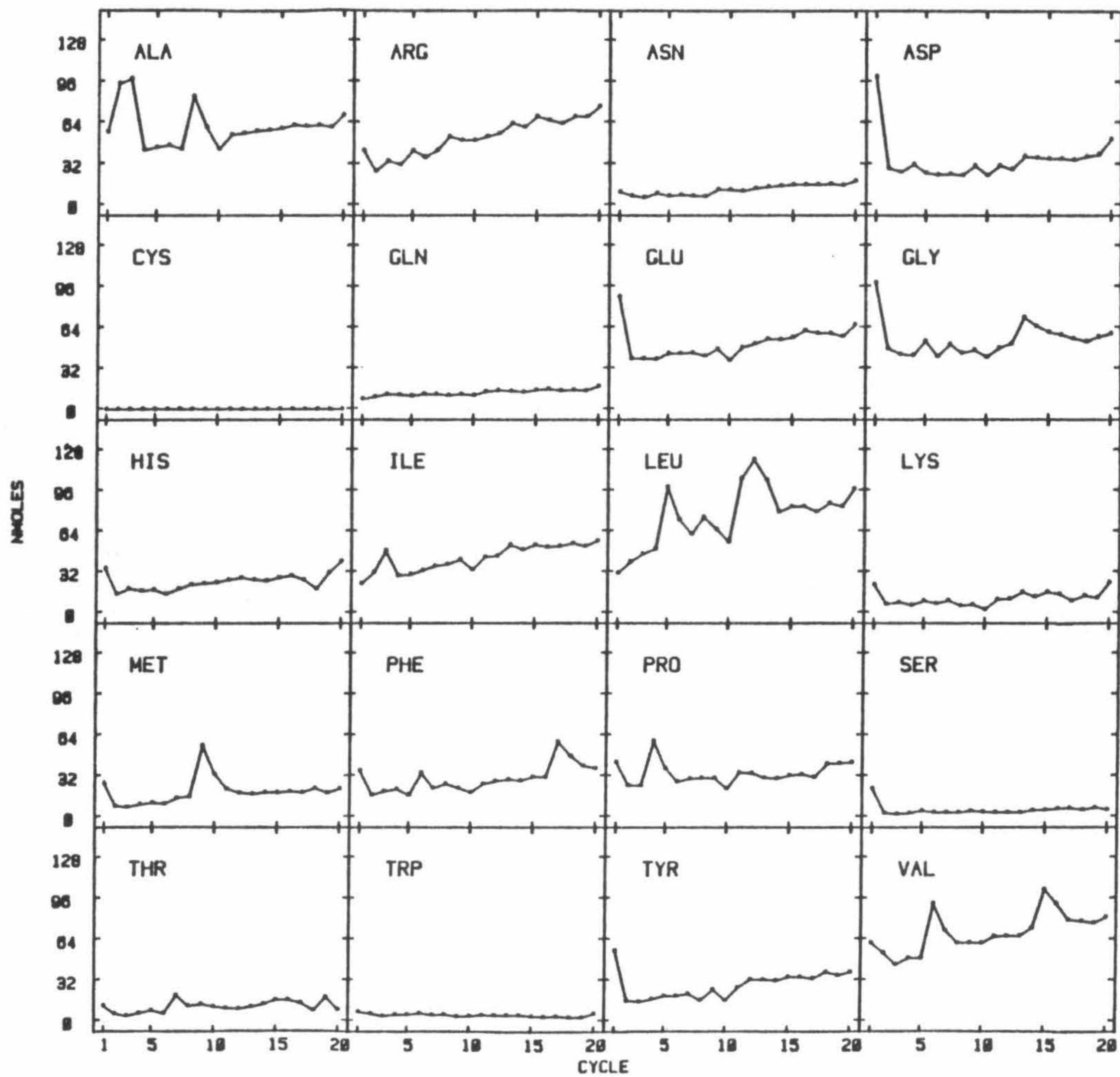

ALA
ARG
ASN
ASP
CYS
GLN
GLU
GLY
HIS
ILE
LEU
LYS
MET
PHE
PRO
SER
THR
TRP
TYR
VAL
NMOLES
128
96
64
32
0
CYCLE
1
5
10
15
20

FIG. 5. The N-terminal sequence of PE2, as determined by Edman degradation. The sequence of PE2 deduced from the yields of PTH derivatives at each cycle, Fig. 4, is labeled "protein." The symbol " " indicates that no residue could be identified, and parentheses indicate some uncertainty in the assignment, when the sequence is based solely on the results of protein sequenation. Also shown is the nucleotide sequence ("mRNA") of a portion of the mRNA which codes for PE2, as well as the translated protein sequence ("translated mRNA") of this RNA sequence, as determined by Rice and Strauss (manuscript in preparation).

1 5

(Ser)-Ala-Ala-Pro-Leu-Val-Thr-Ala-Met-

AAG ACG ACC CCG GAA GGG ACA GAA GAG UGG UCC GCA GCA CCA CUG GUC ACG GCA AUG

-Lys-Thr-Thr-Pro-Glu-Gly-Thr-Glu-Glu-Trp-Ser-Ala-Ala-Pro-Leu-Val-Thr-Ala-Met-

10 15

? -Leu-Leu-Gly- ? -Val- ? -Phe-(Pro)- ? - (protein)

UGU UUG CUC GGA AAU GUG AGC UUC CCA UGC (mRNA)

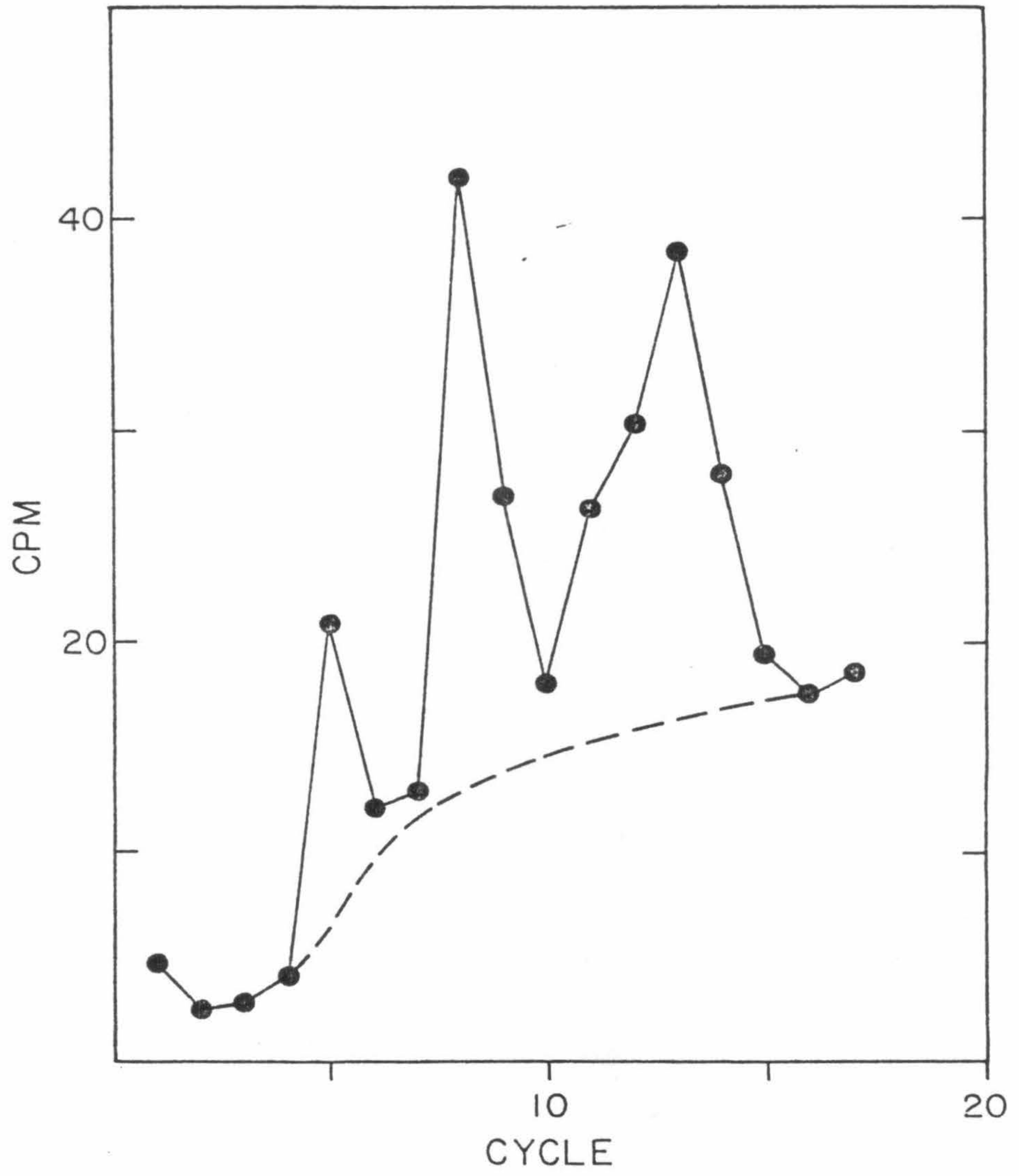
Cys-Leu-Leu-Gly-Asn-Val-Ser-Phe-Pro-Cys- (translated mRNA)

14, and the serine at 16. Our failure to find the cysteine and the serine is not surprising, since the PTH-derivatives of these amino acids are extensively degraded during sequenation, but the asparagine should have been detected if it were unmodified. Since the asparagine of the sequence -Asn-Val-Ser- is a potential glycosylation site (Marshall, 1974), our failure to detect this residue at cycle 14 during Edman degradation strongly implies that it is in fact glycosylated.

An additional result which is apparent from a comparison of the protein sequence and translated mRNA sequence is that the protein sequence which we have obtained represents the primary translation product of the RNA, after cleavage of a single peptide bond to remove the capsid protein sequence. This conclusion is based upon the fact that other workers have isolated the peptide Thr-Thr-Pro-Glu-Gly-Thr-Glu-Glu-Trp from tryptic digests of the Sindbis capsid protein (Boege et al, 1980) and, from Fig. 5, this peptide must be the C-terminus of the capsid protein.

One notable aspect of the sequence data from PE2 presented in Fig. 4 is the low yield of the PTH derivatives, to the extent that less than 10% of the applied protein contained an N-terminus accessible to Edman degradation. This may be due to N-terminal modification of PE2 during purification even though care was taken to minimize this possibility, or it may be that an extensive fraction of PE2 is blocked in vivo. In order to resolve this question, we purified amino acid labeled viral proteins from infected cells by preparative polyacrylamide gel electrophoresis and pooled ^{14}C-leucine labeled PE2 and E2, derived from the same cell lysate and preparative gel, and added ^{14}C-methionine labeled PE2 from a separate preparation. This pool, which was mixed so that PE2 contained twice the CPM per residue as E2 in each labeled position (based on the amino acid compositions of the proteins, unpublished data, and Bell and Strauss [1979]), was subjected to automated Edman degradation, and the mixture of PTH derivatives released at each cycle was counted directly for radioactivity, with the results shown in Fig. 6. The radioactivity at cycles 8 and 13

FIG. 6. Spinning cup sequenator analysis of a mixture of E2 and PE2. The proteins were labeled with ^{14}C-leucine and ^{14}C-methionine so that E2 contained twice the CPM per labeled residue as PE2, and the PTH derivatives released at each cycle were counted directly without fractionation. The dashed line represents an estimate of the background due to internal cleavage of the polypeptide chain during sequenation. Peaks at cycles 8 and 13 are contributed by E2, while peaks at cycles 5, 9, 11, and 12 are from PE2.


40
20
CPM
10
20
CYCLE

was derived from leucine labeled E2 (Bell _et al._, 1979), at cycle 9 from methionine labeled PE2, and the label at cycles 5, 11, and 12 was from leucine labeled PE2. There is also an additional peak of radioactivity at cycle 14, and its presence is due to incomplete cleavage of the Pro_{11}-Tyr_{12} bond of E2 during the appropriate cycle of Edman degradation, which introduces a lag in the following steps of E2 degradation, and the signal found at cycle 14 is the lag of the Leu_{13} signal. Although each labeled position in PE2 contained twice the radioactivity as each labeled E2 position, the peaks derived from E2 are considerably larger than those from PE2. Assuming a repetitive yield of 95%, we calculate that 75% of the E2 was sequencable, compared with only 15% of the PE2. Thus, while PE2 was found to be largely blocked in this experiment also, the E2 which was copurified remained largely unblocked. While these results demonstrate directly that 80-90% of PE2 is blocked _in vivo_, our understanding of the nature of this N-terminal modification is based on results reported elsewhere (Chapter 4, this thesis), in which we have isolated a peptide from the chymotrypsin digestion products of PE2 which is acetylated and which does not appear to contain a free amino group. Taken together, the amino acid composition of this peptide ($ala_{1.9}$, $leu_{1.0}$, $pro_{1.0}$, $ser_{0.8}$) and the N-terminal sequence determined here show that the blocked fraction of PE2 is derived from unblocked PE2 by N-acetylation of the N-terminal serine.

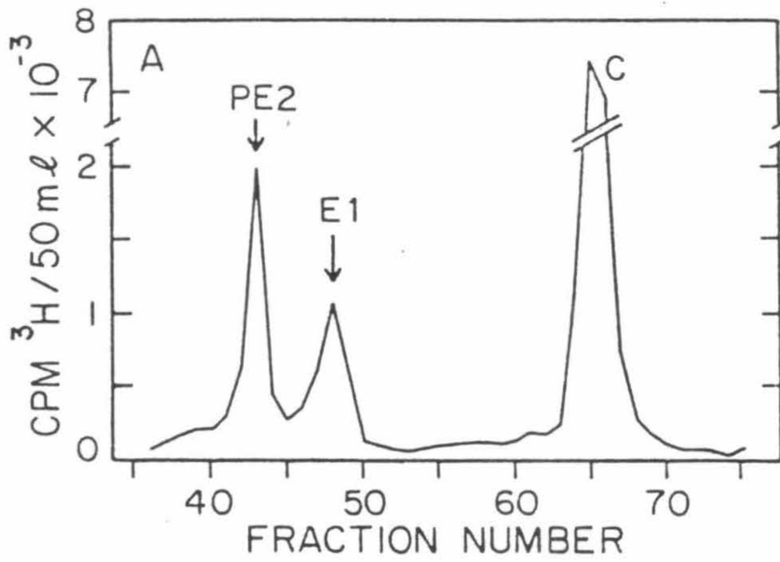
One additional feature of the primary sequence data of Fig. 4 deserves comment, and that is its role in establishing the purity of the PE2 preparation. The background of amino acid derivatives in the early cycles, with the exception of cycle 1, is quite low, especially when one considers that the signal results from sequencing only 10-20% of the PE2 applied. There is a minor sequence, consisting of ?-Val-Ile-Asp-Gly-Phe-?-Leu-, which is the N-terminal sequence of E2 (Bell _et al._, 1978). This E2 sequence is present at about one fourth the level of the unblocked PE2

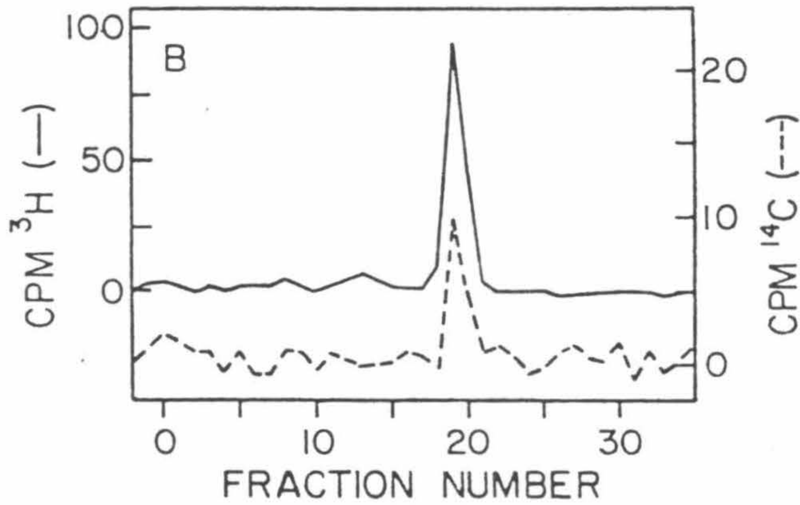
sequence, and its presence may be due to a slight contamination of the preparation with E2 as a result of overloading the preparative gels during purification. Of course, contaminating proteins which also contained a blocked N-terminus would not have been detected.

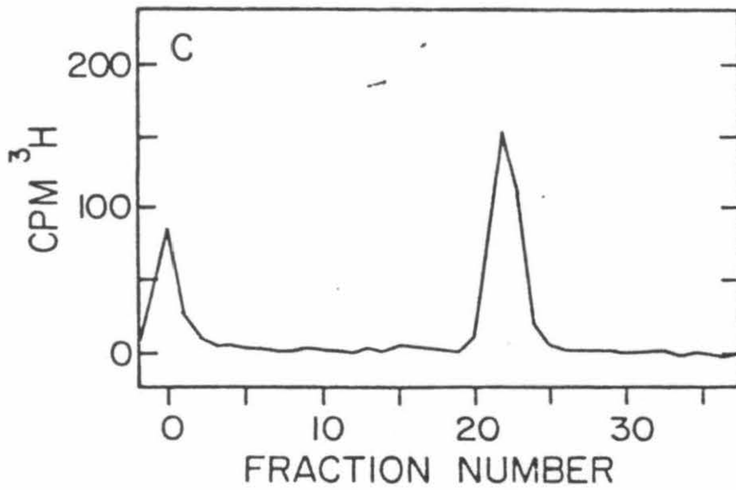
Effect of tunicamycin on the N-terminus of PE2

The finding of the N-terminus of the primary translation product in isolated PE2, albeit largely in an N-acetylated form, is surprising in view of its probable function as a signal sequence in the translocation of PE2 across the membrane of the endoplasmic reticulum during synthesis since such signal sequences are usually rapidly cleaved from the rest of the protein. By comparison with other signal sequences, the presence of carbohydrate in this sequence was also unexpected, and so we decided to see if the N-terminus of PE2 was cleaved in cells treated with tunicamycin, an effective inhibitor of glycosylation (Krag and Robbins, 1977; Leavitt et al., 1977). First, nonglycosylated PE2, labeled with ^{3}H-proline, was purified from tunicamycin-treated infected cells by preparative polyacrylamide gel electrophoresis (Fig. 7a), and PE2 labeled with ^{3}H-acetate (under conditions in which approximately 45% of the incorporated label is in the N^{α}-acetyl group [Chapter 4, this thesis]) and ^{14}C-alanine was similarly purified from untreated cells. We next attempted to isolate the pentapeptide N-acetyl-Ser-Ala-Ala-Pro-Leu, the N-terminal peptide of normal, blocked PE2 (see above), from the nonglycosylated PE2, since if it were found this would show that the N-termini of the normal and of the nonglycosylated PE2 were the same. The isolation procedure, as described elsewhere (Chapter 4, this thesis), consisted of chromatography of an extensive chymotrypsin digest of the protein on Dowex 50W (largely selective for blocked peptides) followed by silica gel chromatography. Figure 7b shows the silica gel chromatogram obtained from the application of the isolation procedure to ^{3}H-acetate and ^{14}C-alanine labeled PE2 from untreated cells; the identification of the acetate-labeled peptide at fraction 18 as the N-

FIG. 7. Effect of tunicamycin on PE2. (a) Preparative polyacrylamide gel electrophoresis of ^{3}H-proline labeled PE2 from tunicamycin treated cells. Electrophoresis was from left to right, and the first 35 and last 22 fractions of the gel were not counted. Peaks are as identified. In a parallel gel, E1 from untreated cells was found at fraction 44. (b) and (c) Silica gel chromatography of PE2 digested with chymotrypsin, and chromatographed on Dowex 50W as described elsewhere (Chapter 4, this thesis). (b) ^{3}H-acetate and ^{14}C-alanine labeled peptides from PE2 purified from untreated cells. (c) ^{3}H-proline labeled peptides from PE2 purified from tunicamycin-treated cells.


A
PE2
E1
C
CPM ^{3}H/50 mℓ × 10^{-3}
8
7
2
1
0
40
50
60
70
FRACTION NUMBER


B
CPM ^{3}H (—)
CPM ^{14}C (---)
100
50
0
20
10
0
0
10
20
30
FRACTION NUMBER


C
CPM ^{3}H
200
100
0
0
10
20
30
FRACTION NUMBER

terminal blocked pentapeptide of normal PE2 is reported elsewhere (Chapter 4, this thesis). As can be seen in Fig. 7c, we were able to isolate this peptide from PE2 from tunicamycin-treated cells also.

DISCUSSION

In order to fully understand the function and processing of precursors of viral structural proteins, it is necessary to purify these precursors for chemical characterization. The approach we have taken to purify PE2, an intracellular precursor to the Sindbis virus envelope protein E2, utilized the extensive purification of E2 provided by its packaging by the infected cell into readily purified virions (Strauss, 1978) which simplified the purification of E2. This E2 was then used for the production of specific antibody, which was in turn used to obtain the precursor from infected cell lysates. This approach should be applicable to the purification of chemically pure precursors to the structural proteins of other viruses, but as far as we know it has not been previously applied to this problem. We were able to use the high resolution of polyacrylamide gel electrophoresis for additional purification due to the extreme sensitivity of the techniques used for subsequent biochemical analysis, which required only small amounts of protein.

An additional aspect of our approach to the purification of PE2 was the choice of the detergent used to solubize PE2. The antibody used was produced in response to SDS-denatured protein, and it was found to be reactive against SDS-denatured components of the virion or SDS-denatured viral proteins in infected cell lysates. The use of SDS at all stages of the purification, except during the binding of PE2 to the immuno-affinity column, was responsible for the high yield of PE2 by eliminating non-specific losses due to sticking of the protein to glassware, for example, and by effecting complete elution of bound PE2 from the immuno-absorbant. We have found

that most of the PE2 which was not recovered in the final preparation could be accounted for in the flow through of the antibody column and much of this could be recovered by subjecting this flow through to a second immuno-absorption (unpublished observations).

Since we quantitated PE2 and E2 during purification with the use of incorporated radiolabel and determined the specific activity of the final preparation, we can calculate the amount of PE2 and E2 in the infected cell at the time of harvest, 7 hr post-infection, or about 5 hr after the initiation of detectable viral protein synthesis (Strauss et al., 1969). At this time, the extent of accumulation of viral proteins is impressive. PE2 constitutes about 0.2% of the total protein in the infected cell, while about 0.35% of the total protein is E2. Labeling experiments indicate that an approximately equal amount of E1 is present, while there is an even greater amount of capsid protein (unpublished observations). We also note that the fraction of E2 and its precursor PE2 which is eventually packaged into virions is rather low. At the time of harvest, we found more than 80 μg of PE2 plus E2 in one roller bottle of infected cells, but after an additional 7 hr of continued synthesis we can recover less than 80 μg of E2 in the virus released from these cells (Bell et al., 1979). Clearly the presence of sufficient amounts of viral structural proteins in the infected cell is not the limiting step in the assembly of the virus particle. These estimates are consistent with previous estimates, obtained at later times in infection, of the viral protein content of infected cells and the extent of packaging of these proteins into virions (Strauss et al., 1969).

A comparison of the partial amino acid sequence of radiolabeled PE2 with that of its counterpart synthesized in vitro in the absence of microsomes led Bonatti and Blobel (1979) to conclude that the signal sequence of PE2 is not cleaved after translocation. We have confirmed this conclusion by showing that isolated PE2 is the primary translation product of this portion of its mRNA, and extended this finding to

include the major, blocked fraction of PE2. We do not know why this blocked fraction was not detected previously, but it may be that PE2 is blocked to a lesser extent in other cells or other virus strains.

The signal sequence of PE2 shares a common feature with other signal sequences, in that they contain a stretch of hydrophobic amino acids (Blobel et al., 1979) (such as is found in the first 19 positions of PE2), although there is otherwise no direct homology. However, in addition to its presence in the completed protein, the signal sequence of PE2 differs from the others in a variety of structural details. Signal sequences of eukaryotes and prokaryotes often contain basic residues at their extreme N-terminal end, and it has been proposed that in bacteria these positively charged groups are involved in the initial binding of the nascent polypeptide to the negatively charged surface of the membrane (Inouye and Halegoria, 1980). The signal sequence of PE2 not only contains no basic residues, 80-90% of PE2 does not even contain a positively charged N^{α}-amino group, this group being involved in an amide bond with an N-acetyl moiety.

The role of N-acetylation of PE2, which makes this the only N-acetylated signal sequence yet found, is unknown. After translocation across the membrane of the endoplasmic reticulum and core glycosylation, PE2 is transported to the Golgi apparatus, where further glycosylation occurs, and is then transported to the plasmalemma. Late in this sequence of events, the cleavage which produces E2 from PE2 occurs (Strauss and Strauss, 1977), possibly just before the protein reaches the cell surface (Rice and Strauss, manuscript in preparation). The N-terminal portion of PE2 is not retained in E2, but it can be recovered in the culture fluid as a glycopeptide 64 amino acids long called E3 (Welch and Sefton, 1979; Rice and Strauss, manuscript in preparation), which has been purified and analyzed in detail (J. T. Mayne, personal communication). Upon Edman degradation, E3 was found to be blocked to a similar extent as PE2 (and the fraction which is sequenceable has the

same N-terminal sequence as the unblocked fraction of PE2) (J. T. Mayne and M. W. Hunkapiller, personal communication). In addition, both blocked and unblocked E3 is glycosylated (J. T. Mayne, personal communication) and there is no potential glycosylation site in the protein other than asparagine$_{14}$ (Rice and Strauss, manuscript in preparation). Thus the overall processing of acetylated and non-acetylated PE2 occurs with equal efficiency.

The recovery of a fraction of PE2 which has escaped N-terminal acetylation probably reflects the manner in which the modified N-terminus of this protein is synthesized. It may be that a reported ribosome-associated acetyltransferase activity is responsible for this modification, although other such enzymatic activities have been found (Pestana and Pitot, 1975). At any rate, of the more than 40 N-acetylated proteins which have been reported (See Narita et al., 1975; Jörnvall, 1975), in all cases this N-terminus is generated in the cytoplasm, although, as is the case with ovalbumin (Palmiter et al., 1978; Lingappa et al., 1979), the acetylated protein may be subsequently excreted. Put another way, we know of no secreted protein in which an N-terminus which is generated by proteolytic cleavage after translocation across the membrane of the endoplasmic reticulum is blocked by N-acetylation, although some such proteins are blocked by the formation of a pyrolidone carboxylate group (see Narita et al., 1975).

We attribute the presence of a minor unblocked fraction of PE2 to the translocation across the membrane of the endoplasmic reticulum, and the separation by this membrane from the acetylating enzyme, of a fraction of the N-termini of PE2 before acetylation occurs. In support of this conclusion is an analysis of the B protein of Sindbis virus. This protein contains the sequences of PE2 and E1 (Schlesinger and Schlesinger, 1973; Bonatti and Blobel, 1979), is not protected from protease digestion by the membrane of microsomal vesicles and is not glycosylated (Sefton and Burge, 1972), and is not processed to yield components of the virion (Duda and Schlesinger,

1975). The B protein thus appears to result from the failure of the N-terminus of PE2 to be inserted through the membrane of the endoplasmic reticulum during synthesis, and this N-terminus is therefore not rapidly separated from the cytoplasm by this membrane. Other workers who obtained N-terminal sequence data from PE2 by Edman degradation found the B protein produced in vivo to be blocked to the extent that no sequence could be obtained (Bonatti and Blobel, 1979).

There is an additional feature of the N-terminus of PE2 which is remarkable, the glycosylation of an asparagine, at position 14, which would appear to be embedded in the signal sequence. Although we have found that this glycosylation is not responsible for the observed lack of cleavage of the signal sequence of PE2, we note that it does result in the attachment of a large hydrophilic group to an otherwise hydrophobic region of the protein, which would be expected to significantly effect the final conformation of the protein since hydrophobic interactions are among the major forces in protein folding (Tanford, 1978). In the case of the protein of certain isolates of vesicular stomatitus virus, the finding that the inhibition of glycosylation also inhibits transport of the protein to the cell surface has been attributed to a change in conformation in the non-glycosylated form of the protein (Gibson et al., 1979). It may well be that the similar inhibition of transport of the Sindbis virus glycoproteins when glycosylation is prevented (Leavitt et al., 1977) is due at least in part to a change in the conformation of PE2 when its N-terminus lacks carbohydrate.

The uniqueness of the N-terminus of PE2, when compared with other signal sequences, makes it unlikely that the transmembrane translocation of secreted and membrane proteins in general can be completely described by a few simple rules. This is also apparent from the way in which ovalbumen is translocated across the membrane prior to secretion, for in this case the signal sequence has been found to be internal in the polypeptide chain (Lingappa et al., 1979). An even more interesting example of the extent of the complexity of membrane protein synthesis is provided by

studies in a prokaryotic system, the cotranslational insertion of the λ receptor into the outer membrane of *E. coli* (Emr *et al.*, 1980). This protein contains a 25 amino acid signal sequence which is structually similar to eukaryotic signal sequences and which is cleaved during translocation of the protein. However, additional information in the sequence of the λ receptor is required for the insertion of a cytoplasmic protein into the membrane when the cytoplasmic protein is fused, by genetic techniques, to a portion of the N-terminus of the membrane protein. The availability of the signal sequence of PE2 in E3, although in a modified form may be of use in the further characterization of the biochemical events involved in the transmembrane translocation of proteins.

ACKNOWLEDGEMENTS

We wish to thank L. E. Hood for his advice and assistance and E. Lenches for preparing the chick embryo fibroblasts. This investigation was supported by grants BM06965 and AI10793 from the National Institutes of Health and by grant PCM77-26728 from the National Science Foundation. J.R.B. was supported by training grant GM00086 from the National Institutes of Health.

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 composition of the structural proteins of Sindbis virus. Virology **97**, 287-294.

Bethell, G. S., Ayers, J. S., Hancock, W. S., and Hearn, M. T. W. (1979). A novel method of activation of cross-linked agarose with 1,1'-carbonyldiimidazole which gives a matrix for affinity chromatography devoid of additional charged groups. J. Biol. Chem. **254**, 2572-2574.

Blobel, G., and Dobberstein, B. (1975). Transfer of proteins across membranes II. Reconstitution of functional rough microsome by heterologous components. J. Cell Biol. **67**, 852-862.

Blobel, G., and Sabatini, D. D. (1970). Controlled proteolysis of nascent polypeptides in rat liver cell fractions. J. Cell Biol. **45**, 130-145.

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.

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.

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.

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 **283**, 433-438.

Duda,E., and Schlesinger, M. J. (1975). Alterations in Sindbis and viral envelope proteins by treating BHK cells with glucosamine. J. Virol. **15**, 416-419.

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.

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., and Soderlund, 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.

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

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

Helenius, A., and Simons, K. (1975). Solubilization of membranes by detergents. Biochim. Biophys. Acta **415**, 29-79.

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 increased sensitivity. Science **207**, 523-525.

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

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.

Jörnvall, H. (1975). Acetylation of protein N-terminal amino groups. Structural observations on α-amino acetylated proteins. J. Theor. Biol. **55**, 1-12.

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.

Krag, S. S., and Robbins, P. W. (1977). Sindbis envelope proteins as endogenous asceptors in reactions of guanosine diphosphate-[^{14}C] mannose with preparations of infected chicken embryo fibroblasts. J. Biol. Chem. **252**, 2621-2629.

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

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

Lingappa, V. R., Katz, F. N., Lodish, H. F., and Blobel, G. (1978). A signal sequence for the insertion of a transmembrane glycoprotein: similarities to the signals of secretory proteins in primary structure and function. J. Biol. Chem. **253**, 8667-8670.

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

Malkin, L. I., and Rich, A. (1967). Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. J. Mol. Biol. **26**, 329-346.

Marshall, R. D. (1974). The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. Biochem. Soc. Symp. **40**, 17-26.

Narita, K., Matsuo, H., and Nakajima, T. (1975). End group determination. In "Protein Sequence Determination", 2nd edition (S. B. Needleman, ed.), pp. 30-103. Springer-Verlag, New York.

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.

Palmiter, R. D., Gagnon, J., and Walsh, K. A. (1978). Ovalbumin: a secreted protein without a transient hydrophobic leader sequence. Proc. Natl. Acad. Sci. USA **75**, 94-98.

Pestana, A., and Pitot, H. C. (1975). Acetylation of nascent polypeptide chains in rat liver polyribosomes in vivo and in vitro. Biochemistry **14**, 1404-1412.

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., and Schlesinger, S. (1973). Large-molecular-weight precursors of Sindbis virus proteins. J. Virol. **11**, 1013-1016.

Sefton, B. M., and Burge, B. W. (1973). Biosynthesis of the Sindbis virus carbohydrates. J. Virol. **12**, 1366-1374.

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

Strauss, J. H., Burge, B. W., and Darnell, J. E. (1969). Sindbis virus infection of chick and hamster cells: synthesis of virus-specific proteins. Virology **37**, 367-376.

Strauss, E. G., and Strauss, J. H. (1977). Togaviruses In "The Molecular Biology of Animal Viruses: (D. Mayak, ed.), Vol. 1, pp-111-116, Dekker, New York.

Tanford, C. (1978). The hydrophobic effect and the organization of living matter. Science **200**, 1012-1018.

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.

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.