Proc. Natl. Acad. Sci. USA
Vol. 77, No. 11, pp. 6376–6380, November 1980
Biochemistry

The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino-terminal region

(nucleotide sequencing/cDNA/molecular cloning/protein–RNA interactions)

H. GAROFF, A.-M. FRISCHAUF, K. SIMONS, H. LEHRACH, AND H. DELIUS

European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 10.2209, 69 Heidelberg, Federal Republic of Germany

Communicated by John C. Kendrew, July 21, 1980

ABSTRACT **The amino acid sequence of the capsid (C) protein was deduced from the nucleotide sequence of the *C* gene. This part of the viral 42S RNA genome was transcribed into double-stranded cDNA. The cDNA was cloned in the *Escherichia coli* χ1776–pBR322 host–vector system and then the base sequence was determined with the technique described by Maxam and Gilbert. The amino acid sequence of the C protein shows a clustering of basic amino acids and prolines within the first 110 amino acids.**

Semliki Forest virus (SFV) is a simple membrane virus of the alphavirus group. It has been used extensively as a model system to study the structure and assembly of cellular membranes (1). The virus particle consists of an icosahedral nucleocapsid surrounded by a membrane. The nucleocapsid is a complex of about 240 capsid proteins (C protein, M_r = 30,000) (2, 3) and a RNA molecule (42S), the viral genome (4). The membrane consists of a lipid bilayer with about 240 external glycoprotein spikes (5). Each spike contains three different glycopolypeptides: E1 (M_r = 49,000), E2 (M_r = 52,000), and E3 (M_r = 10,000) (6, 7). The E2 polypeptide spans the membrane; there are about 30 amino acid residues present on the internal side of the viral membrane (8, 9).

The virus enters the host cell by absorptive endocytosis (10). Inside the lysosomes of the cell, the low pH probably triggers a fusion of the viral membrane with the lysosomal membrane (10, 11). This allows the nucleocapsid to enter the cell cytoplasm, where the viral genome is uncoated so that it can act as a mRNA for synthesis of polymerase molecules. The viral RNA polymerase synthesizes new 42S RNA molecules and smaller 26S RNA molecules. The latter molecule is homologous to the 3′ end of the viral genome (12) and functions as a mRNA for the SFV structural proteins, which are translated from a single initiation site (13). The C protein is made first. As soon as it is completed it is cleaved from the growing polypeptide chain, and the ribosomes continue to read off the membrane proteins in the order E3, E2, and E1 (8, 14). The membrane proteins are cotranslationally translocated across the membrane of the endoplasmic reticulum and transported to the plasma membrane (15, 16).

The assembly of the nucleocapsid in the cell cytoplasm is not understood. Newly synthesized capsid proteins are known to be associated with the large subunit of the ribosome before they complex with the 42S RNA into nucleocapsids (17). The final step in SFV assembly, budding, takes place at the cell surface (18). The nucleocapsid binds to the cytoplasmic aspect of the plasma membrane which folds around the nucleocapsid. In the budding process the virus-spike glycoproteins are included into the viral membrane, whereas the host proteins are excluded from the virus. This specificity can probably be explained by the formation of bonds between the C protein in the nucleocapsid and the spanning segments of the spike glycoproteins (9).

A more detailed understanding of SFV structure and assembly at the molecular level is difficult without the knowledge of the amino acid sequences of the structural proteins. We report here the primary structure of the C protein.

MATERIALS AND METHODS

Materials. Oligo(dT)-cellulose (T3) was from Collaborative Research (Waltham, MA). All isotopes were purchased from Radiochemical Centre (Amersham, England). Restriction endonuclease *Xho* I and *Ava* I were from P-L Biochemicals. *Cla* I, *Hin*dIII, *Hin*dII, *Alu* I, *Pst* I, and *Bam*HI were from Boehringer. *Hpa* II, *Hha* I, *Hin*fI, *Hae* III, *Sau* I, *Taq* I, and *Eco*RI were gifts from V. Pirrotta. Reverse transcriptase was generously provided by J. Beard. DNA polymerase was a gift from W. McClure. S1 nuclease and T4 polynucleotide kinase were prepared as described (see ref. 19).

Isolation of RNA. The 42S RNA was isolated from purified virus particles. SFV (5 mg of protein) was dissociated in 5 ml of 10 mM Tris, pH 7.4/10 mM NaCl/1.5 mM $MgCl_2$/1% Na$DodSO_4$. The clear solution was extracted twice with redistilled phenol/chloroform/isoamyl alcohol, 2.4:2.4:0.1 (vol/vol). The RNA in the waterphase was precipitated with ethanol and dissolved in H_2O (1 mg/ml).

The 26S RNA was isolated from infected BHK-21 cells. Thirty Falcon bottles (75 cm^2) with a monolayer of BHK-21 cells were infected with SFV at about 100 plaque-forming units per cell (20). The cells were cooled on ice 6 hr after infection, and a cytoplasmic extract was prepared as described (21). This extract (≈25 ml) was made 2% in Na$DodSO_4$ and then diluted to 100 ml with 10 mM Tris, pH 7.4/50 mM NaCl. After extracting twice with phenol/chloroform/isoamyl alcohol, 48:48:2 (vol/vol), the nucleic acids in the water phase were precipitated and taken up in 10 mM Tris, pH 7.4/1 mM EDTA/0.1% Na$DodSO_4$. From this preparation poly(A)-containing RNA molecules were isolated by chromatography on oligo(dT)-cellulose (22) and fractionated further on a 10–30% (wt/vol) sucrose gradient (23). The RNA in the 26S fraction was collected, precipitated, and dissolved in H_2O (≈1 mg/ml). All RNA samples were stored at −70°C.

Preparation of Double-Stranded cDNA. About 10 μg of 42S RNA was used as template for cDNA synthesis. The procedure to prepare double-stranded (ds) cDNA has been described (19). For colony hybridizations (see below) single-stranded (ss) cDNA that contained one ^{32}P-labeled nucleotide was used.

Molecular Cloning. The ds cDNA was inserted into the

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

Abbreviations: SFV, Semliki Forest virus; C, capsid; ds, double stranded; ss, single-stranded; kb, kilobase; bp, base pair.

Pst I site of the plasmid pBR322 by using the oligo(dG-dC)-tailing procedure described by Röwekamp and Firtel (24). *Escherichia coli* χ1776 was transformed with the hybrid plasmid (25) in a P3 physical containment laboratory at the European Molecular Biology Laboratory. After growth the colonies were hybridized against ^{32}P-labeled ss cDNA from 42S RNA to detect chimeric plasmids (26). Several strongly hybridizing colonies were found. These were screened for length of the hybrid plasmid by using the miniscreening technique of Barnes (27). The clones containing the longest hybrid plasmids were then grown in 1-liter cultures. Hybrid plasmids were purified from these clones either by chromatography on hydroxyapatite (28) or by equilibrium centrifugation in dye/CsCl gradients. *E. coli* tRNA, which still contaminated the plasmid fraction, was removed by gel filtration. All plasmid preparations were stored at −30°C in H_2O (1 mg/ml).

This work was done in accordance with the German guidelines for recombinant DNA research.

Characterization of Hybrid Plasmids. A preliminary characterization of the insert in the hybrid plasmid was made with electron microscopy. The insert was cleaved from the vector with *Pst* I and isolated by electrophoresis on an agarose gel (see below). Purified inserts were then denatured and hybridized to 42S or 26S RNA in a buffer containing 80% (vol/vol) formamide (29). The nucleic acids were treated with gene 32 protein of the T4 phage and adsorbed to mica sheets (30). The length of the insert molecule and its location (that is, its distance from the 3′ end of the RNA molecule) were measured. Values in base pairs were obtained by comparison to the DNA of phage PM 2 as an internal standard [10 kilobases (kb)] or to *in vitro* transcripts of phage T7. A molecular weight of 4.2×10^6 was found for the 42S RNA. This corresponds to 12.7 kb.

Mapping with restriction endonucleases was done as described by Smith and Birnstiel (32). The digestion conditions for the various restriction endonucleases were as follows: 10 mM Tris, pH 7.5/10 mM $MgCl_2$/10 mM dithiothreitol was used with *Xho* I, *Ava* I, *Hin*fI, *Hpa* II, *Hae* III, *Hha* I, *Sau* I, and *Bam*HI. The $MgCl_2$ concentration was lowered to 2 mM for digestion with *Taq* I. NaCl (50 mM) was included in the buffer for digestions with *Hin*dII, *Hin*dIII, *Pst* I, *Alu* I, and *Cla* I; 100 mM NaCl, was included in digestions with *Eco*RI. The incubation temperature was 70°C for *Taq* I, 30°C for *Pst* I, and 37°C for the rest. Partial digestion products of the end-labeled DNA molecule were obtained by transferring one fourth (2.5 μl) of the reaction volume at 5, 15, 30, and 60 min to 10 μl of 40 mM Tris, pH 7.8/5 mM sodium acetate/0.1 M EDTA/2.5% glycerol/ 0.025% bromphenol blue. The amount of enzyme added was calculated to give a nearly complete digestion in 60 min.

Nucleotide Sequence Determinations. DNA fragments were end-labeled with ^{32}P and the sequence was determined by using base-specific chemical cleavages as described by Maxam and Gilbert (33). The polynucleotide kinase reaction was used to label 5′ termini with ^{32}P. Recessive 3′ ends of DNA fragments were labeled in reactions with DNA polymerase and [^{32}P]deoxynucleoside triphosphates (34). End-labeled fragments were isolated by agarose gel electrophoresis and subsequent electroelution into hydroxyapatite as described by Tabak and Flavell (35). The two strands of the labeled DNA fragments were dissociated in 0.3 M NaOH (10 min at 37°C) and separated by electrophoresis in 6% (wt/vol) polyacrylamide gels (20 × 40 cm) (36). The front plate was cooled to 2°C. After electrophoresis at 300 V for 10 hr, the bands were located by autoradiography, appropriate zones were cut from the gel, and the DNA was eluted by shaking the gel pieces at 37°C for 12 hr in 0.5 ml of 40 mM Tris, pH 7.8/5 mM sodium acetate/1 mM EDTA.

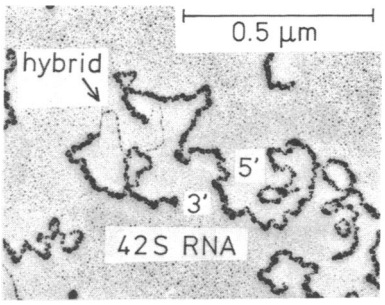

FIG. 1. Electron micrograph of 42S RNA hybridized to the SFV cDNA insert of pBR SFV 3 isolated after cleavage with *Pst* I. The ds hybrid region appears as a thin thread in comparison with the ss region complexed with gene *32* protein.

End-labeled DNA samples that had gone through five base-specific cleavage reactions (G, G+A, C+T, C, and C+A) were analyzed on sequencing gels essentially as described (37). The thickness of the gel was reduced to 0.2 mm. In some runs the temperature of the gel was kept at 70°C by using a plate heated with water instead of one of the normal glass plates. Before autoradiography the gel was fixed in 10% acetic acid for 10 min and dried in an oven at 70°C. In order to make handling of the thin gel easy, it was covalently bound to one of the glass plates by treating this with a silane mixture before casting the gel (unpublished data).

The nucleotide sequences read from the gels were overlapped by using a computer program that will be described elsewhere.

RESULTS

Characterization of a Clone Containing the C Gene. Electron microscopic measurements on hybrids between the recombinant plasmid DNA and SFV RNA suggested that one of the plasmids, pBR SFV 3 contained the entire *C* protein gene. Length measurements of the insert molecule when hybridized to 42S RNA (Fig. 1) indicated that it contained about 1600 bases and that its 5′ end was located some 2800 bases from the 3′ end of the 42S RNA molecule. Hybridization against 26S RNA (Fig.

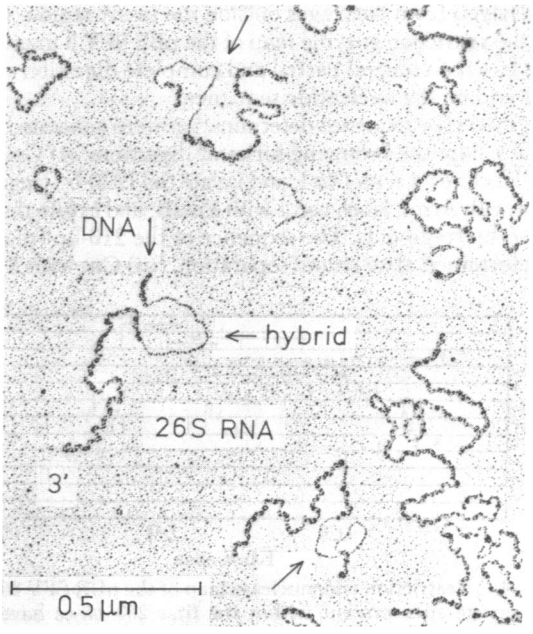

FIG. 2. Electron micrograph of 26S RNA hybridized to the SFV cDNA insert of pBRSFV 3.

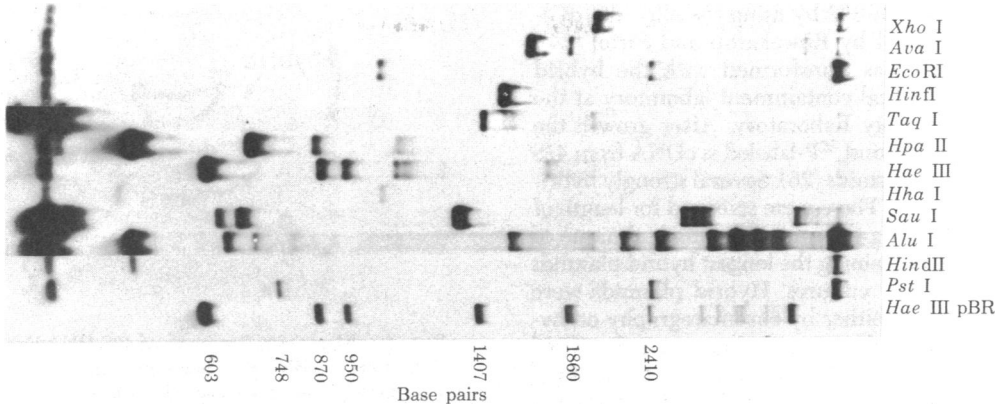

FIG. 3. Analysis of end-labeled partial digestion products of pBR SFV 3 on a 1.4% agarose gel (20 × 40 cm; 1.5 mm thick). The enzymes used are indicated at the top. The small and the large *Pst* I fragments of pBR SFV 3 mark the limits of the insert region. Approximate lengths in base pairs (bp) were calculated by using the *Hae* III partial digestion products of end-labeled pBR322 (right lane) as standard lengths (31).

2) showed that the length of the hybrid region was about 250 bases shorter than the hybrid region in 42S RNA. There was a short and a long stretch of unhybridized nucleic acid on the sides. The shorter stretch was about 250 bases long and the longer stretch was about 2800 bases long. These results suggest that the inserted DNA of pBR SFV 3 includes about 1400 bases of the 5′ end of the 26S RNA and in addition some 250 bases adjacent to the 26S RNA region on the 42S RNA. The coding region for the C protein should be totally included in this insert (13, 14).

The cleavage sites for various restriction endonucleases were mapped on the pBR SFV 3 insert using the procedure as described (32). The hybrid plasmid was linearized with *Cla* I, end-labeled using the polynucleotide kinase reaction, and recleaved with *Hin*dIII. *Hin*dIII cleaved the DNA six bases to the right of the *Cla* I site on the vector portion of pBR SFV 3 (38). This resulted in a DNA molecule with a single label 750 bases away from the insert region. Fig. 3 shows the partial digestion pattern of end-labeled pBR SFV 3 when treated with several different restriction endonucleases and analyzed on a vertical agarose gel. The products migrating faster than the small *Pst* I fragment or slower than the large *Pst* I fragment were derived from cleavages outside the insert region. Fig. 4 shows the restriction enzyme map of the pBR SFV 3 insert. The map is based on several partial and complete digestion experiments and on the nucleotide sequence.

Fragments for sequence determination were generated from pBR SFV 3 by the following series of digestions. (*i*) Cut with *Xho* I, label, recut with *Cla* I, and isolate the 1920-bp fragment. (*ii*) Cut with *Ava* I, label, recut with *Eco*RI, and isolate the 596- and 216-bp fragments. The sequence of the 216-bp fragment was determined after strand separation. (*iii*) Cut with *Eco*RI, label, recut with *Xho* I, and isolate the 947-bp fragment. (*iv*) Cut with *Hin*fI, label, recut with *Cla* I, and isolate the 1447-bp fragment or recut with *Xho* I and isolate the 473-bp fragment. (*v*) Cut with *Taq* I, label, recut with *Hin*dII, and isolate the 881-bp fragment. The *Cla* I and the *Hin*dII sites were located on the vector part of the hybrid plasmid. The base sequence from the left *Eco*RI site to the *Xho*I site is shown in Fig. 5.

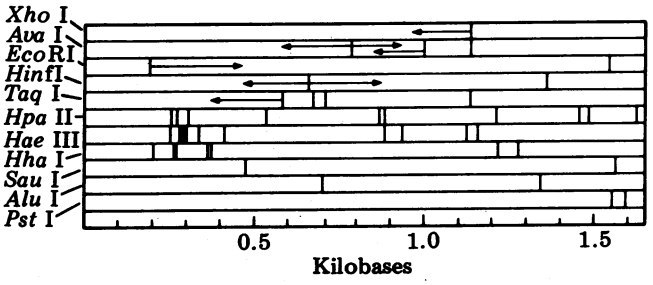

FIG. 4. Restriction endonuclease map of the pBR SFV 3 insert. The cleavage sites present within the first 200 bases have been omitted. The arrows indicate stretches of DNA in which the sequence of both strands have been determined.

Amino Acid Sequence of the C Protein. Translation of the base sequence into amino acid sequences in all three reading frames showed that only one of the reading frames was open throughout. The second reading frame contained 10 nonsense codons, and the third reading frame contained 12. This suggested that reading in frame one gives the correct amino acid sequence. In the beginning of this sequence (Fig. 5) there was a Met-Asn dipeptide. This must represent the amino terminus of the C protein because the same peptide had been found to be labeled with formyl [^{35}S]methionine from initiator tRNA in C protein made *in vitro* from the 26S RNA (13). The carboxy-terminal tryptophan (39) was 267 amino acids further along the chain just before the amino terminus of the E3 protein

Table 1. Amino acid composition of the C protein*

	From sequence	Experimental†
Lys	36	33.9
His	7	7.0
Arg	15	13.8
Asp	12	21.5
Asn	10	
Thr	16	15.7
Ser	10	11.1
Glu	11	28.7
Gln	15	
Pro	23	22.1
Gly	20	24.5
Ala	23	23.5
1/2 Cys	4	4.4
Val	19	17.2
Met	8	7.0
Ile	11	10.5
Leu	9	10.2
Tyr	7	6.7
Phe	6	6.4
Trp	5	2.7

* M_r = 29,828.
† Ref. 2.

```
AGA ATT CTG ATT ATA GCG CAC TAT TAT AGC ACC ATG AAT TAC ATC CCT ACG CAA ACG TTT TAC GGC CGC CGC TGG  (14)
                                            MET ASN TYR ILE PRO THR GLN THR PHE TYR GLY ARG ARG TRP

CGC CCG CGC CCG GCG GCC CGT CCT TGC CGC TTG CAG GCC ACT CCG GTG GCT CCC GTC GTC CCC GAC TTC CAG GCC
ARG PRO ARG PRO ALA ALA ARG PRO TRP PRO LEU GLN ALA THR PRO VAL ALA PRO VAL VAL PRO ASP PHE GLN ALA  (39)

CAG CAG ATC CAG CAA CTC ATC AGC GCC GTA AAT GCG CTG ACA ATG AGA CAG AAC GCA ATT GCT CCT GCT AGG CCT
GLN GLN MET GLN GLN LEU ILE SER ALA VAL ASN ALA LEU THR MET ARG GLN ASN ALA ILE ALA PRO ALA ARG PRO  (64)

CCC AAA CCA AAG AAG AAG AAG ACA ACC AAA CCA AAG CCG AAA ACG CAG CCC AAG AAG ATC AAC GGA AAA ACG CAG
PRO LYS PRO LYS LYS LYS LYS THR THR LYS PRO LYS PRO LYS THR GLN PRO LYS LYS ILE ASN GLY LYS THR GLN  (89)

CAG CAA AAG AAG AAA GAC AAG CAA GCC GAC AAG AAG AAG AAG AAA CCC GGA AAA AGA GAA AGA ATG TGC ATG AAG
GLN GLN LYS LYS LYS ASP LYS GLN ALA ASP LYS LYS LYS LYS LYS PRO GLY LYS ARG GLU ARG MET CYS MET LYS  (114)

ATT GAA AAT GAC TGT ATC TTC GAA GTC AAA CAC GAA GGA AAG GTC ACT GGG TAC GCC TGC CTG GTG GGC GAC AAA
ILE GLU ASN ASP CYS ILE PHE GLU VAL LYS HIS GLU GLY LYS VAL THR GLY TYR ALA CYS LEU VAL GLY ASP LYS  (139)

GTC ATG AAA CCT GCC CAC GTG AAA GGA GTC ATC GAC AAC GCG GAC CTC GCA AAG CTA GCT TTC AAG AAA TCG AGC
VAL MET LYS PRO ALA HIS VAL LYS GLY VAL ILE ASP ASN ALA ASP LEU ALA LYS LEU ALA PHE LYS LYS SER SER  (164)

AAG TAT GAC CTT GAG TGT GCC CAG ATA CCA GTT CAC ATG AGG TCG GAT GCC TCA AAG TAC ACG CAT GAG AAG CCC
LYS TYR ASP LEU GLU CYS ALA GLN ILE PRO VAL HIS MET ARG SER ASP ALA SER LYS TYR THR HIS GLU LYS PRO  (189)

GAG GGA CAC TAT AAC TGG CAC CAC GGG GCT GTT CAG TAC AGC GGA GGT AGG TTC ACT ATA CCG ACA GGA GCG GGC
GLU GLY HIS TYR ASN TRP HIS HIS GLY ALA VAL GLN TYR SER GLY GLY ARG PHE THR ILE PRO THR GLY ALA GLY  (214)

AAA CCG GGA GAC AGT GGC CGG CCC ATC TTT GAC AAC AAG GGG AGG GTA GTC GCT ATC GTC CTG GGC GGG GCC AAC
LYS PRO GLY ASP SER GLY ARG PRO ILE PHE ASP ASN LYS GLY ARG VAL VAL ALA ILE VAL LEU GLY GLY ALA ASN  (239)

GAG GGC TCA CGC ACA GCA CTG TCG GTG GTC ACC TGG AAC AAA GAT ATG GTG ACT AGA GTG ACC CCC GAG GGG TCC
GLU GLY SER ARG THR ALA LEU SER VAL VAL THR TRP ASN LYS ASP MET VAL THR ARG VAL THR PRO GLU GLY SER  (264)

GAA GAG TGG TCC GCC CCG CTG ATT ACT GCC ATG TGT GTC CTT GCC AAT GCT ACC TTC CCG TGC TTC CAG CCC CCG
GLU GLU TRP

TGT GTA CCT TGC TGC TAT GAA AAC AAC GCA GAG GCC ACA CTA CGG ATG C
```

FIG. 5. The nucleotide sequence from the left *Eco*RI site to the *Xho* I site (see Fig. 4) on the pBR SFV 3 insert, with the deduced amino acid sequence for the C protein beneath.

(unpublished data). Table 1 shows the amino acid composition and M_r of the C protein as calculated from the amino acid sequence. The values are very close to those found experimentally (2). After this work was completed, Boege *et al.* (40) reported the amino acid sequence and composition data of several tryptic peptides derived from the SFV C protein. Their data are accommodated within our nucleotide sequence.

The sequence of the 267 amino acids in the C protein shows a striking clustering of basic amino acids and proline within the first 110 residues of the protein (Fig. 5). This segment includes 21 lysine residues, 9 arginine, and 16 proline. Only three aspartic acids and one glutamic acid are present within this region. The first basic cluster is an arginine pair at positions 12 and 13. This is followed shortly by three Arg-Pro sequences. There is a Arg-Pro-Pro-Lys-Pro-Lys-Lys-Lys-Lys sequence starting at position 63. Close to this sequence there is a Lys-Pro-Lys-Pro-Lys sequence starting at position 74 and a Pro-Lys-Lys sequence starting at position 81. Three lysines are found at positions 92–94 and five lysines and one proline at positions 100–105. The carboxy terminus is acidic. Three glutamic acid residues are found within the six last residues. The rest of the polypeptide chain shows a relatively even distribution of basic and acidic amino acids.

DISCUSSION

The C protein plays a fundamental role in the assembly of the SFV particle. At least three interactions are of importance: interactions between C protein subunits, between C proteins and the 42S RNA, and between C proteins and the virus spike glycoproteins. Newly made C proteins bind rapidly to 42S RNA molecules in the cytoplasm of infected cells. A complex of one RNA molecule with its complement of C proteins folds to form the icosahedral nucleocapsid. Empty capsids have not been detected in the infected cell. Apparently the C proteins cannot self-associate to form the capsid shell. Binding to the RNA seems necessary for nucleocapsid assembly. After the nucleocapsid has been assembled, it diffuses to the plasma membrane, where it probably becomes bound to the cytoplasmic domains of the virus spike glycoproteins that are spanning the plasma membrane (9). The surface-bound nucleocapsid then acts as a template for the budding process, ensuring incorporation of the virus spike glycoproteins into the budding segment of the plasma membrane. Each C protein is postulated to have one binding site for the virus spike glycoproteins, and when all the binding sites on the C proteins of the nucleocapsid are filled, the virus particle is released into the extracellular medium.

Before a virus can infect a host cell and start a new round of virus replication, most of the C protein interactions established during assembly of the virus particle have to be disrupted. This presumably takes place after the virus has entered the lysosomes by adsorptive endocytosis (10, 11). It is possible that the acid pH of the lysosomes not only causes fusion of the SFV membrane with the lysosomal membrane but also destabilizes the protein–protein and the protein–RNA interactions of the C protein. The SFV nucleocapsid undergoes a conformational change upon treatment with slightly acidic buffer (pH 6.0) (41). Whether this structural change induces uncoating is not known.

The striking feature of the primary structure of the SFV C protein is the clustering of basic amino acid residues (lysine and arginine) with proline within the first 110 amino acid residues from the amino terminus of the protein. We suggest that these

clusters are involved in RNA binding in the SFV nucleocapsids. The abundant basic side chains offer charges for salt bridges with phosphate groups in the RNA. Clusters of basic amino acid residues and proline have also been found in histone H1 from sea urchin (42) and to a lesser extent in other histone proteins (43). It is interesting in this context that histone–DNA complexes dissociate at low pH, the HI histone being the first removed (50). A Pro-Lys-Arg-Lys sequence and a Pro-Arg-Pro sequence are also present in the amino-terminal region of the polyoma virus VP1 protein (44). A Pro-Lys-Lys-Pro-Lys sequence is found in the corresponding region of the SV40 virus VP1 protein (45). VP1 is the major protein component in the icosahedral nucleocapsid of these viruses. In the carboxy-terminal region of the core antigen from hepatitis B virus, the sequence Ser-Pro-$(Arg)_3$ occurs three times, and there is a double repeat of the sequence Ser-Pro-$(Arg)_4$-Ser-Gln (46).

There is evidence that suggests that RNA–protein interactions are of major importance in stabilizing the SFV nucleocapsid. The RNA in the nucleocapsid is sensitive to low concentrations of RNase. Such treatment leads to considerable shrinkage of the nucleocapsid and to disruption of the nucleocapsid structure at higher RNase concentrations (47). The nucleocapsid is also sensitive to concentrations of $NaDodSO_4$ that are considerably lower than those normally needed to disrupt protein–protein interactions (48). A number of plant viruses, of which cucumber mosaic virus is the prototype, show similar sensitivity to RNase and $NaDodSO_4$ (49). It has been suggested that these viruses are stabilized mainly by RNA–protein interactions. It will be interesting to see whether the primary structures of the coat proteins from these viruses also possess extensive clusters of basic amino acids and prolines.

We would like to thank Evelyn Kiko and Hilkka Virta for excellent technical assistance, Vince Pirrotta for stimulating discussions, Ken Murray for a critical reading of the manuscript, and Annie Biais and Wendy Moses for typing the manuscript.

1. Simons, K., Garoff, H., Helenius, A. & Ziemiecki, A. (1978) in *Frontiers in Physicochemical Biology*, ed. Pullman, B. (Academic, New York), pp. 387–407.
2. Simons, K. & Kääriäinen, L. (1970) *Biochem. Biophys. Res. Commun.* **5**, 981–988.
3. Laine, R., Söderlund, H. & Renkonen, O. (1973) *Intervirology* **1**, 110–118.
4. Simmons, D. T. & Strauss, J. H. (1972) *J. Mol. Biol.* **71**, 599–613.
5. Bonsdorff, C.-H. & Harrison, S. C. (1975) *J. Virol.* **16**, 141–145.
6. Garoff, H., Simons, K. & Renkonen, O. (1974) *Virology* **61**, 493–504.
7. Ziemiecki, A. & Garoff, H. (1978) *J. Mol. Biol.* **122**, 259–269.
8. Garoff, H. & Söderlund, H. (1978) *J. Mol. Biol.* **124**, 535–549.
9. Garoff, H. & Simons, K. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 3988–3992.
10. Helenius, A., Kartenbeck, J., Simons, K. & Fries, E. (1980) *J. Cell Biol.* **84**, 404–420.
11. White J. & Helenius, A. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3273–3277.
12. Kennedy, S. I. T. (1976) *J. Mol. Biol.* **108**, 491–511.
13. Glanville, N., Ranki, M., Morser, J., Kääriäinen, L. & Smith, A. E. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3059–3063.
14. Clegg, J. C. S. (1975) *Nature (London) New Biol.* **254**, 454–455.
15. Garoff, H., Simons, K. & Dobberstein, B. (1978) *J. Mol. Biol.* **124**, 587–600.
16. Ziemiecki, A., Garoff, H. & Simons, K. (1980) *J. Gen. Virol.*, in press.
17. Ulmanen, I., Söderlund, H. & Kääriäinen, L. (1976) *J. Virol.* **20**, 203–210.
18. Acheson, N. H. & Tamm, I. (1967) *Virology* **32**, 123–143.
19. Lehrach, H., Frischauf, A.-M., Hanahan, D., Wozney, J., Fuller, F. & Boedtker, H. (1979) *Biochemistry* **18**, 3146–3152.
20. Kääriäinen, L., Simons, K. & von Bonsdorff, C.-H. (1969) *Ann. Med. Exp. Biol. Fenn.* **47**, 235–248.
21. Glanville, N., Morser, J., Uomala, P. & Kääriäinen, L. (1976) *Eur. J. Biochem.* **64**, 167–175.
22. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
23. Dobberstein, B., Garoff, H., Warren, G. & Robinson, P. J. (1979) *Cell* **17**, 759–769.
24. Röwekamp, W. & Firtel, R. (1980) *Dev. Biol.*, in press.
25. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, E., Naber, S. P., Chick, W. L. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3727–3731.
26. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
27. Barnes, W. M. (1977) *Science* **195**, 393–394.
28. Colman, A., Byers, M. J., Primrose, S. B. & Lyons, A. (1978) *Eur. J. Biochem.* **91**, 303–310.
29. Chow, L. T., Roberts, J. M., Lewis, J. B. & Broker, T. R. (1977) *Cell* **11**, 819–836.
30. Priess, H., Koller, B., Hess, B. & Delius, H. (1980) *Mol. Gen. Genet.* **178**, 27–34.
31. Sutcliffe, J. G. (1978) *Nucleic Acids Res.* **5**, 2721.
32. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3**, 2387–2398.
33. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
34. Soeda, E., Kimura, G. & Miura, K.-L. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 162–166.
35. Tabak, H. F. & Flavell, R. A. (1978) *Nucleic Acids Res.* **5**, 2321–2332.
36. Hayward, G. S. (1972) *Virology* **49**, 342–344.
37. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
38. Sutcliffe, J. G. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3737–3741.
39. Kalkkinen, N. (1980) *FEBS Lett.* **115**, 163–166.
40. Boege, U., Wengler, G., Wengler, G. & Wittmann-Liebold, B. (1980) *Virology* **103**, 178–190.
41. Söderlund, H., Kääriäinen, L., Bonsdorff, C.-H. & Weckstrom, P. (1972) *Virology* **47**, 753–760.
42. Strickland, W. N., Strickland, M., Brandt, W. F., von Holt, C., Lehmann, A. & Wittmann-Liebold, B. (1980) *Eur. J. Biochem.* **104**, 567–578.
43. Dayhoff, M. (1978) *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3.
44. Soeda, E., Arrand, J. R. & Griffin, B. E. (1980) *J. Virol.* **33**, 619–630.
45. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Hereweghe, J., Volckaert, G. & Ysebaert, M. (1978) *Nature (London)* **273**, 113–120.
46. Pasek, M., Goto, T., Gilbert, W., Zink, B., Schaller, H., Mackay, P., Leadbetter, G. & Murray, K. (1979) *Nature (London)* **282**, 575–579.
47. Söderlund, H., von Bonsdorff, C.-H. & Ulmanen, I. (1979) *J. Gen. Virol.* **45**, 15–26.
48. Becker, R., Helenius, A. & Simons, K. (1975) *Biochemistry* **14**, 1835–1841.
49. Kaper, J. M. (1973) *Virology* **55**, 299–304.
50. Murray, K. (1969) *J. Mol. Biol.* **39**, 125–144.