PROTEINS: Structure, Function, and Genetics 27:345–359 (1997)

Structure of Semliki Forest Virus Core Protein

Hok-Kin Choi,[1] Guoguang Lu,[1] Sukyeong Lee,[1] Gerd Wengler,[2] and Michael G. Rossmann[1]*
[1]*Department of Biological Sciences, Purdue University, West Lafayette, Indiana, 47907*
[2]*Institut für Virologie, Justus-Liebig-Universität Giessen (FB18), Frankfurter Str. 107, D-35392, Giessen, Germany*

ABSTRACT **Alphaviruses are enveloped, insect-borne viruses, which contain a positive-sense RNA genome. The protein capsid is surrounded by a lipid membrane, which is penetrated by glycoprotein spikes. The structure of the Sindbis virus (SINV) (the type virus) core protein (SCP) was previously determined and found to have a chymotrypsin-like structure. SCP is a serine proteinase which cleaves itself from a polyprotein. Semliki Forest virus (SFV) is among the most distantly related alphaviruses to SINV. Similar to SCP, autocatalysis is inhibited in SFCP after cleavage of the polyprotein by leaving the carboxy-terminal tryptophan in the specificity pocket. The structures of two different crystal forms (I and II) of SFV core protein (SFCP) have been determined to 3.0 Å and 3.3 Å resolution, respectively. The SFCP monomer backbone structure is very similar to that of SCP. The dimeric association between monomers, A and B, found in two different crystal forms of SCP is also present in both crystal forms of SFCP. However, a third monomer, C, occurs in SFCP crystal form I. While monomers A and B make a tail-to-tail dimer contact, monomers B and C make a head-to-head dimer contact. A hydrophobic pocket on the surface of the capsid protein, the proposed site of binding of the E2 glycoprotein, has large conformational differences with respect to SCP and, in contrast to SCP, is found devoid of bound peptide. In particular, Tyr184 is pointing out of the hydrophobic pocket in SFCP, whereas the equivalent tyrosine in SCP is pointing into the pocket. The conformation of Tyr184, found in SFCP, is consistent with its availability for iodination, as observed in the homologous SINV cores. This suggests, by comparison with SCP, that E2 binding to cores causes major conformational changes, including the burial of Tyr184, which would stabilize the intact virus on budding from an infected cell. The head-to-tail contacts found in the pentameric and hexameric associations within the virion utilize the same monomer surface regions as found in the crystalline dimer interfaces. Proteins 27:345–359, 1997.**

Key words: alphavirus structure; Semliki Forest virus capsid protein; autocatalysis; capsid assembly; conformational changes

INTRODUCTION

Semliki Forest virus (SFV) is one of 26 members of the alphavirus family.[1] These are arthropod-borne, enveloped viruses, which can cause encephalitis, fever, arthritis, and rash in mammals.[2] The virus particle consists of an RNA genome packaged in an icosahedral nucleocapsid with $T = 4$ quasi-symmetry. It is surrounded by a membrane[3,4] penetrated by 80 glycoprotein "spikes," also arranged in a $T = 4$ surface lattice.[3,4] Each spike consists of three copies of E1, E2, and (in SFV, but not in Sindbis virus [SINV]) E3. These spikes contain the recognition site for the host cell receptor.[5]

The structural proteins of alphaviruses are translated from a subgenomic mRNA into a polyprotein. The capsid protein is at the amino terminus and cleaves itself from the polyprotein by a *cis* autocatalytic process.[6,7] The crystal structure of Sindbis virus core protein (SCP)[8,9] showed that the amino-terminal domain was either missing or disordered. The C-terminal domain consists of 151 amino acid residues and is folded into a chymotrypsin-like structure with *Ser215*, *His141*, and *Asp163* forming the catalytic triad.[8–10] (Throughout this paper, Semliki Forest core protein (SFCP) residue numbers will be written in bold characters, while SCP numbers will be italicized, unless otherwise stated.) The proteinase is inhibited after performing the first catalytic cleavage, because the carboxy-terminal tryptophan remains in the specificity pocket near the active center.[8,11] The amino acid sequence Gly-Asp-Ser-Gly, which is conserved among all the chymotrypsin-like serine proteinases and contains the catalytically essential serine, is present in all the alphavirus core proteins for which the amino acid sequence is

Dr. Choi's present address is DuPont Merck Pharmaceutical Company, Research and Development, Experimental Station, P.O. Box 80336, Wilmington, DE 19880-0336.
Dr. Lu's present address is Molecular Structural Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institute, 17177 Stockholm, Sweden.
*Correspondence to: Dr. Michael G. Rossmann, Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392.
Received 9 September 1996; accepted 11 September 1996.

known.[12–19] Viral proteinases, including SCP, differentiate themselves from chymotrypsin in having shorter insertions between the β strands forming each of the two β barrels. The similarity in the fold of SCP to that of chymotrypsin suggests that the virus acquired a proteinase from a host at some time in the past.

Cryoelectron microscopy (cryo-EM) image reconstructions and computer fitting of the known structure of SCP[8,9] have shown that the carboxy-terminal domain of alphavirus core proteins form pentamers and hexamers on the surface of the internal nucleocapsid.[4] Monomers are associated primarily by binding the loop residues **122–132** in one subunit with loop residues **185–195** and **222–232** in the neighboring subunit. Recognition of the spike proteins during the budding process[15,20–23] is probably mediated by binding of the carboxy-terminal residues of E2 into a hydrophobic pocket, consisting in part of **Tyr184** and **Trp251,** on the surface of the capsid protein.[24,25]

SFCP consists of 267 amino acids[16] homologous to SCP. The basic amino-terminal domains, 118 residues in SFCP and 113 in SCP, have considerably less sequence similarity than do the carboxy-terminal domains (Fig. 1). While the SFCP amino-terminal domain is particularly rich in lysine, the corresponding SCP domain has no lysine until amino acid *71*. The carboxy-terminal domains have an overall amino acid identity of 65%, but the SFCP residues **170** to **267** have an even higher identity of 76% to SCP. The differences between the sequences of SFCP and SCP and between the properties of SFV and SINV are the biggest among the 26 members of the alphavirus genus.

We report here the structure of SFCP and compare it with SCP, the only other known alphavirus capsid protein structure.

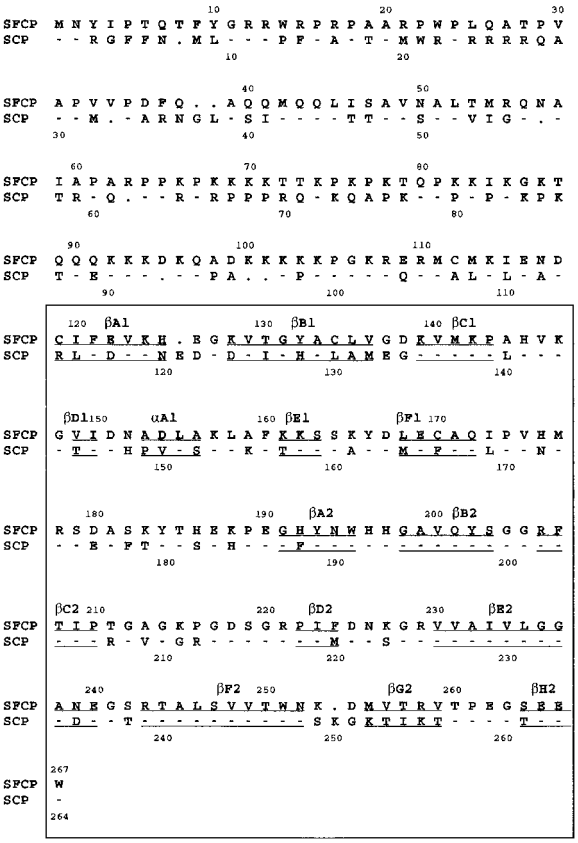


Fig. 1. The sequence alignment of the SIN and SFV core proteins. A dash represents an amino acid that is identical in both proteins. A dot represents a deletion. The rectangular enclosure is the carboxy-terminal domain seen in the crystallographic structure determination. Secondary structural elements are underlined.

MATERIALS AND METHODS

Propagation and Protein Extraction

The techniques used for SFV propagation and protein extraction followed those developed for SINV by Boege et al.[19] SFV (prototype strain) was propagated in baby hamster kidney cells (BHK21). The virus was incubated with the nonionic detergent NP40, permitting extractions of the nucleocapsids on equilibrium sucrose density gradients. The nucleocapsids were then disrupted with acetic acid and urea, and the SFCP recovered after stepwise removal of urea and acetic acid by dialysis and lyophilization.

Crystallization and Heavy Atom Derivative Preparation

Crystals were grown in hanging drops at room temperature. The reservoir solution contained 6–10% (w/v) polyethylene glycol (PEG) 8000 in 200–300 mM Tris buffer at pH 8.3. A stock protein solution was made with a concentration of 10 mg/ml in water. The initial drops consisted of 5 µl of reservoir solution, plus 5 µl of protein solution. Small crystals appeared after several months. Addition of 10 mM K_2HgI_4 helped to form bigger and better diffracting "type I" crystals. The type I crystals were used for both native and heavy atom derivative data collection. Because the crystals were unstable in artificial mother liquor, the $KAuCl_4$ and $KAu(CN)_2$ heavy atom derivatives were prepared by slowly transferring 2 µl of the 10 mM heavy atom solution into hanging drops. "Type II" crystals were formed by crushing type I crystals and using the resultant seeds in conditions identical to those used for obtaining the initial seed crystals.

Data Collection, Processing, and Packing Considerations

X-ray diffraction data (Table I) were collected for native, $KAuCl_4$ and $KAu(CN)_2$ derivatized type I crystals and for native type II crystals on a Siemens area detector system mounted on an Elliott GX20 rotating anode generator operated at 35 kV and 40 mA. All crystals were maintained at about 12°C. The

TABLE I. Diffraction Data Collection

A. *Type I native*					
Resolution limit (Å)	10.0	6.30	4.50	3.60	3.00
No. of reflections	262	834	1717	2091	1933
Completeness (%)	93.6	97.4	89.2	72.2	44.8
R_{merge} (%)*	4.2	6.1	6.8	7.7	11.8
Oscillation (angle/frame)	0.15°				
Oscillation (time/frame)	8 min				
Number of crystals used	2				
B. *Type I $KAuCl_4$*					
Resolution limit (Å)	10.0	6.30	4.50		
No. of reflections	236	718	1221		
Completeness (%)	84.3	83.9	63.5		
R_{merge} (%)*	4.0	9.0	13.2		
R_{diff} (%)*	27.6	28.5	22.8		
Oscillation (angle/frame)	0.30°				
Oscillation (time/frame)	15 min				
Number of crystals used	1				
C. *Type I $KAu(CN)_2$*					
Resolution limit (Å)	10.0	6.30	4.50		
No. of reflections	226	630	922		
Completeness (%)	80.7	73.6	47.9		
R_{merge} (%)*	4.6	9.0	11.8		
R_{diff} (%)*	28.7	26.1	21.3		
Oscillation (angle/frame)	0.30°				
Oscillation (time/frame)	15 min				
Number of crystals used	1				
D. *Type II Native*					
Resolution limit (Å)	10.0	6.30	4.50	3.60	3.30
No. of reflections	218	628	1501	1837	1286
Completeness (%)	94.4	99.5	99.1	94.5	80.7
R_{merge} (%)*	3.4	4.8	5.3	7.1	9.3
Oscillation (angle/frame)	0.15°				
Oscillation (time/frame)	5 min				
Number of crystals used	2				

*
$$R_{merge} = \frac{\sum_h \sum_i |I_{hi} - I_h|}{\sum_h \sum_i I_{hi}} \times 100; \quad R_{diff} = \frac{\sum |(|F_{PH}| - |F_P|)|}{\sum |F_P|} \times 100.$$

indexing, integration and processing were carried out using the XDS program.[26]

The type I crystals have a $P2_1$ space group with a volume per unit molecular weight (V_M)[27] of 2.18 or 1.45 Å^3/D, assuming a molecular weight of about 30,000 per SFCP monomer and 2 or 3 monomers per asymmetric unit, respectively.[7,28] However, if the initial 118 residues are omitted from the molecular weight calculation, the V_M value is 4.01 and 2.68 Å^3/Da, respectively. A comparison of the V_M values for the different crystal forms shows that type I crystals might have 2 or 3 monomers per crystal asymmetric unit (Table II). Type II crystals were also of space group $P2_1$, but with a smaller unit cell. The V_M value suggested that these crystals had 2 monomers per asymmetric unit (Table II).

Structure Determination of Type I Crystals

A self-rotation function was computed with the program GLRF[29,30] for data between 15.0 and 6.0 Å resolution, using a 40.0 Å radius of integration. The $\kappa = 180°$ plane showed that the highest peak (other than the origin) had polar angles $\phi = 106.5°$ and $\psi = 59.0°$ (see ref. 29 for the definition of the polar angles).

A polyalanine model of the SCP dimer was placed in a $P2$ cell with $a = 80$, $b = 150$, $c = 80$ Å, and $\alpha = \beta = \gamma = 90.0°$ for the purpose of calculating a cross-rotation function. The twofold axis was made parallel to the *c*-axis and the longest dimension of the molecule was along the *b*-axis. The cell was large enough to avoid overlap of the self Pattersons. The cross-rotation function was calculated for data between 15 and 6 Å resolution, with a 40 Å radius of integration. A peak was found that was 5.0σ above the mean value of the map. The next highest peak was 3.9σ above the mean. This result showed that the dimer axis of the model was oriented in the SFCP unit cell with polar angles $\phi = 107.1°$ and $\psi = 59.0°$, consistent with the self-rotation function results. A

TABLE II. SCP and SFCP Crystal Forms

Viral protein	Crystal type	Space group	Cell dimensions	Z*	Apparent V_M† (Å³/D)	Effective V_M‡ (Å³/D)
SCP	1	$P222_1$	a = 37.0 Å b = 293.2 Å c = 83.3 Å	5?	1.5	2.6
SCP	2	$P4_32_12$	a = 57.0 Å b = 57.0 Å c = 109.8 Å	1	1.5	2.6
SCP	3	$P2_1$	a = 38.8 Å b = 79.7 Å c = 60.8 Å β = 102.2°	2	1.6	2.7
SCP	4	$P1$	a = 29.0 Å b = 56.5 Å c = 60.8 Å α = 93.1° β = 96.7° γ = 94.9°	2	—	2.8
SFCP	I	$P2_1$	a = 39.1 Å b = 176.2 Å c = 40.2 Å β = 110.4°	2 3	2.2 1.5	4.0 2.7
SFCP	II	$P2_1$	a = 53.7 Å b = 48.6 Å c = 70.4 Å β = 105.8°	2	1.5	2.7

*Z is the number of SCP or SFCP monomers in the crystallographic asymmetric unit.
†The apparent V_M is based on the molecular weight of the protein used for crystallization.
‡The effective V_M is based on the molecular weight of only the ordered part of the protein structure.

translation function was calculated using data between 8 and 4 Å resolution.[31] Because the y coordinate is indeterminate in space group $P2_1$, the origin of the rotated dimer model was arbitrarily placed at $y = 0.0$. The top peak was 7.0σ above the mean, with the next highest peak at 2.8σ.

Phases were generated between 30.0 and 3.0 Å resolution by using the polyalanine backbone structure of the model, given the SCP dimer orientation and position in the SFCP unit cell. The phases of the calculated structure amplitudes were combined with the observed structure factors and used to compute an electron density map of SFCP at 3.0 Å resolution. An atomic mask was generated from the atomic coordinates of the ordered part of the SCP dimer molecule. The resultant mask occupied 52% of the unit cell volume. The electron density map was then averaged within the mask and solvent flattened outside the mask. The modified map was Fourier back-transformed, and the process was repeated for six iterations.[32] The overall correlation coefficient* between observed and calculated structure amplitudes improved from 0.55 to 0.78. However, the packing of the atomic mask in the unit cell showed that there was a complete layer in the crystal where there was no protein-protein contact.

The phases generated by molecular replacement averaging were used to interpret $KAuCl_4$ and $KAu(CN)_2$ heavy atom difference maps (Table III). The $KAuCl_4$ difference electron density map had four positive and three negative peaks well above background. The four positive peaks were consistent with the noncrystallographic symmetry, with one site near **His192** and the other near **His196,** consistent with anticipated Au substitutions. Two of the three negative peaks, in noncrystallographically equivalent sites, were close to **Cys119** and **Cys134.** Therefore, these negative sites were interpreted as Hg substitutions in the "native" data set, for which the crystals had been grown in the presence of K_2HgI_4. This would imply that the $KAuCl_4$ derivative probably contained lower occupancies of Hg than the native data set. The third negative peak was in the empty solvent region and could not be associated with a noncrystallographically equivalent site. A difference electron density map of the $KAu(CN)_2$ compound showed only three negative peaks corresponding to the Hg sites in the native map. Multiple isomorphous replacement (MIR) phasing and molecu-

*The correlation coefficient is defined as

$$\frac{\Sigma(\langle F_o\rangle - F_o)(\langle F_c\rangle - F_c)}{[\Sigma(\langle F_o\rangle - F_o)^2\,\Sigma(\langle F_c\rangle - F_c)^2]^{1/2}}$$

where F_o are the observed structure amplitudes and F_c are the structure amplitudes of the inverted modified electron density map.

TABLE III. Refined Heavy Atom Parameters for Type I Crystals

	Monomer	Site	x*	y*	z*	$Z^{\dagger}$	Atom
$KAuCl_4$ derivative	A	1	0.3684	−0.0830	0.0814	26.1	Au
		2	0.1038	−0.0146	0.0020	30.7	Au
		3	0.4555	−0.0742	0.4998	−40.3	Hg
	B	1	0.0376	0.0142	−0.2302	16.6	Au
		2	0.3623	−0.0338	0.0450	18.3	Au
		3	0.1652	0.0971	0.0029	−24.0	Hg
	C	3	−0.1568	0.2603	0.1157	−28.2	Hg
$KAu(CN)_2$ derivative	A	4	0.4509	−0.0730	0.5088	−51.8	Hg
	B	4	0.1630	−0.0931	0.0059	−20.3	Hg
	C	4	−0.1545	0.2586	0.1092	−50.2	Hg

*x, y, z are the fractional coordinates of the sites.
†Z is the relative occupancy.

lar replacement phasing were combined in order to interpret density in the vacant parts of the cell, but no easy interpretation was immediately possible.

The SFCP dimer structure was built with respect to the available electron density map. The structure was tentatively refined with X-PLOR.[33] This model was used to compute a new electron density map at 4.5 Å resolution. The MIR density within the mask for the dimer was then replaced by the model electron density. Care was taken to ensure that the mean MIR density outside the mask was equal to the mean density produced by the SFCP model. The result was Fourier back-transformed, and the calculated phases were combined with the observed amplitudes to compute a new electron density map. For the next cycle, the density within the mask was again replaced by the model SFCP dimer density. Six such cycles were completed. Subsequently, the resolution was extended by about one reciprocal lattice unit at a time in a succession of 15 steps, which brought the resolution to 3.0 Å. Visual inspection of the density outside the mask tentatively suggested the presence of a third monomer.

The density corresponding to one of the monomers within the dimer was then flattened and the resultant map was Fourier back-transformed. The resultant calculated structure factors were compared with those of an SFCP monomer by a fast-rotation function,[34] in which the largest peak corresponded to the unflattened monomer. The height of the second largest peak above background suggested that there was a third monomer in the asymmetric unit. A subsequent translational function[34] demonstrated that the empty space within the unit cell was occupied by a third monomer (C) of the SFCP carboxy-terminal domain. This structure was related to monomer B by means of a pseudo noncrystallographic twofold axis. The orientation of this dimer axis corresponded to the second largest peak in the original self-rotation function. Furthermore, the Hg site previously found in the "empty" part of the cell (see above) was seen to bind to the equivalent cysteine residues of monomer C.

The structure was refined by the program X-PLOR, using the Engh and Huber[35] idealized conformational parameters. The refined structure, consisting of the three independent monomers A, B, and C, was restrained by the noncrystallographic dimer axis with the exception of certain residues. These residues were found to have different conformations as judged by omit maps. Each monomer contained the amino acids from residue **119** to the carboxy-terminal tryptophan at residue **267,** amounting to 3429 nonhydrogen atoms. Included also were the three Hg atoms, but solvent molecules were excluded. The S—Hg—S angle was taken as 180° and the S—Hg bond length was taken as 2.46 Å. The R factor dropped to 18.8% by using an overall B factor. Subsequently, strongly restraint-independent temperature factors were used, resulting in a final R factor of 15.7% (see Table IV for refinement details). Well-defined continuous electron density was observed in the final $(2F_o - F_c)$ map for all the main chain atoms and most of the side chain atoms. However, the side chains of the surface residues, **Trp267A*** and Trp267C had poorer definition. Residue **Lys253** has a different conformation in all three monomers and is located in a loop which makes lattice contacts in monomers A and B. **Lys253B** lies in the generously allowed region of the Ramachandran diagram. Coordinates have been deposited with the Brookhaven Protein Data Bank (identification code 1VCP).

Structure Determination of Type II Crystals

The unrefined model of the SFCP AB dimer was used as a search model to solve the type II crystal structure. The program AMoRe[34] was used to calculate a Crowther fast-rotation function[36] to ascertain whether the type II crystals contained the same AB dimer as found in SCP and type I SFCP structures. The cross-rotation function was calculated for data

*SFCP sequence numbers are followed by an A, B or C when necessary to show the monomer identity.

TABLE IV. Statistics of Structure Refinement

	Type I crystal*	Type II crystal
Data used for the refinement (Å)	6.0–3.0	6.5–3.3
Number of nonhydrogen atoms	3429	2284
Number of reflections	5452	4501
R-value	0.157	0.181
r.m.s. bond lengths (Å)	0.013	0.014
r.m.s. bond angles (°)	1.7	1.8
r.m.s. dihedral angles (°)	27.5	28.1
r.m.s. improper angles (°)	1.5	1.6
Main chain dihedral angles (%)		
in most favored regions	85.1	86.6
in additional allowed regions	14.6	13.5
in generously allowed regions	0.3	none
in disallowed regions	none	none
B_{mean} (Å^2) of all atoms	12.9	18.6
B_{mean} (Å^2) of main chain atoms	12.1	17.6
B_{mean} (Å^2) of side chain atoms	13.7	19.7
r.m.s. *B* cross-bond (Å^2)†	1.4	0.7
r.m.s. *B* cross-bond angles‡	2.0	1.0

*Residues not restrained by the noncrystallographic symmetry: **128, 136–139, 141, 162, 165, 172, 186, 215, 238, 240, 252255, 260.** r.m.s. distance between restrained equivalent atoms in type I crystal form: A to B 0.19 Å, B to C 0.20 Å, C to A 0.19 Å.

†The value of r.m.s. *B* cross-bonds was calculated as $(\Sigma_j\Sigma_i(B_{ij} - B_{jmean})^2/2 \cdot N_{bond})^{1/2}$, where B_{ij} represents the *B* -factor of the *i*th atom in the *j*th bond, B_{jmean} represents the average *B*-factor of the *j*th bond, and N_{bond} represents the number of bonds in the structure.

‡The value of r.m.s. *B* cross-bond angles was calculated as $(\Sigma_j\Sigma_i(B_{ij} - B_{jmean})^2/3 \cdot N_{angle})^{1/2}$, where B_{ij} represents the *B* -factor of the *i*th atom in the *j*th bond, B_{jmean} represents the average *B*-factor of the *j*th bond, and N_{angle} represents the number of bonds in the structure.

between 10 and 4 Å resolution with a 30 Å radius of integration. The largest peak, at 12 σ above background, was twice as high as the next largest peak. The resultant orientation was employed in the AMoRe translation function with data between 8 and 4 Å resolution. The top peak, at 13 σ, was twice as high as the next largest peak. The *R* factor was 41.5% on placing the unrefined SFCP dimer (found from the type I crystals) into the $P2_1$ type II crystal unit cell. The structure was then refined using X-PLOR in a manner similar to that described for the type I crystal in which the two molecules were restrained by the noncrystallographic symmetry (see Table IV for the refinement details). Coordinates have been deposited with the Brookhaven Protein Data Bank (identification code 1VCQ).

RESULTS AND DISCUSSION

Molecular Interactions

In at least two different crystal forms (types 2 and 3, Table II), SCP forms a dimer in which two monomers are related by a twofold axis. This dimer has only a small surface area of contact (~400 Å^2) and is not found in intact virions.[4]

The structures of two different crystal forms (types I and II) of SFCP were determined (Table II). In crystal form I, there were three monomers (A, B, and C), and in crystal form II, there were two monomers (A and B) per asymmetric unit, giving a total of 5 crystallographically independent structures of SFCP. The "AB" dimers observed in two crystal forms of SCP[8,11] were also observed in both forms of SFCP. In the type I SFCP crystal structure, the AB dimer is complexed with a third monomer, C, making an "ABC" trimer. The interaction between the A and B monomers is between the carboxy-terminal subdomains (tail-to-tail), whereas the interaction between the B and C monomers is between the amino-terminal subdomains (head-to-head) (Fig. 2 and Table V). In contrast, the interactions within the pentamers and hexamers that occur in alphavirus capsids are head-to-tail[4] (Fig. 3 and Table V). The carboxy-terminal subdomain contacts that exist in the frequently found tail-to-tail AB dimer are also utilized in the head-to-tail contacts within the virion. The "BC" dimer contact surface, found in the type I SFCP crystal structure, borders on the head-to-tail contact surface in the virions (Fig. 3).

In the type I SFCP crystal structure, the orientation of the AB and BC dimer axes differs by 15.0°, and the shortest distance between these dimer axes is 30 Å (Fig. 2). In the type II SFCP crystal structure, the AB dimer in one unit cell is in contact with an A′B′ dimer in the next unit cell, such that the BA′ contact between dimers in adjacent cells resembles the BC contact of type I crystals. However, the crystallographic symmetry requires that the BA′ dimer axis must be parallel (rather than inclined by 15°) to the AB dimer axis (Fig. 2).

The dimer interface between the A and B monomers (Fig. 4a) covers only 374 Å^2 in SFCP or 409 Å^2 in SCP, whereas between monomers B and C (Fig. 4b) the contact area is somewhat more extensive with 495 Å^2. However, while there are 6 hydrogen bonds in the AB interface of SFCP, there are only 2 hydrogen bonds in the BC interface of SFCP. The absence of a BC dimer in any of the SCP structures is undoubtedly due to the presence of *Lys155*, which is an alanine in SFCP. This lysine would be close to the BC dimer axis and, thus, would create a large repulsive force in SCP.

The AB monomer interactions of SFCP are similar to that of SCP, except *Phe188*, buried in the interface of SCP, is replaced by **His192** in SFCP. This causes the side chain of each **Asn194** to make an extra hydrogen bond with His 192 in the opposing monomer (Fig. 4a).

Structure Description

The first 118 amino acids of SFCP are not visible in the present electron density maps. The carboxy-terminal 149 residues are folded into two "Greek key" β barrel subdomains (β strands A_1, B_1, C_1, D_1,

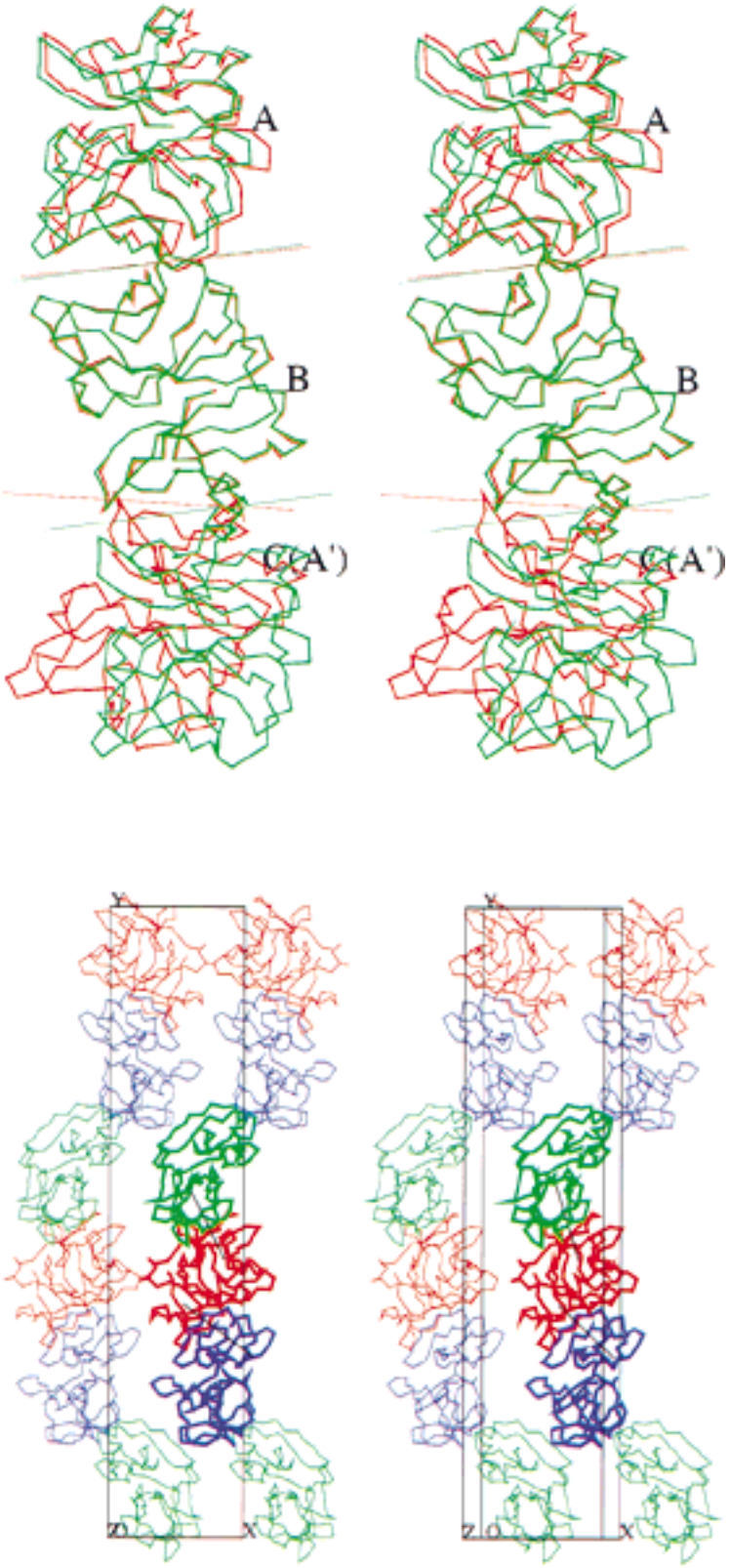


Fig. 2. Crystal packing arrangement of SFCP monomers showing dimer axes. **Top:** Red is crystal type I with independent monomers A, B, and C within one asymmetric unit. Green is crystal type II with independent monomers A and B. A′ is a crystallographic repeat. Red and green straight lines indicate the dimer axes in type I and II crystals, respectively. **Bottom:** Packing of the trimer A (green), B (red), and C (purple) in the $P2_1$ type I unit cell. The bold monomers represent one asymmetric unit. The dimer axes are black.

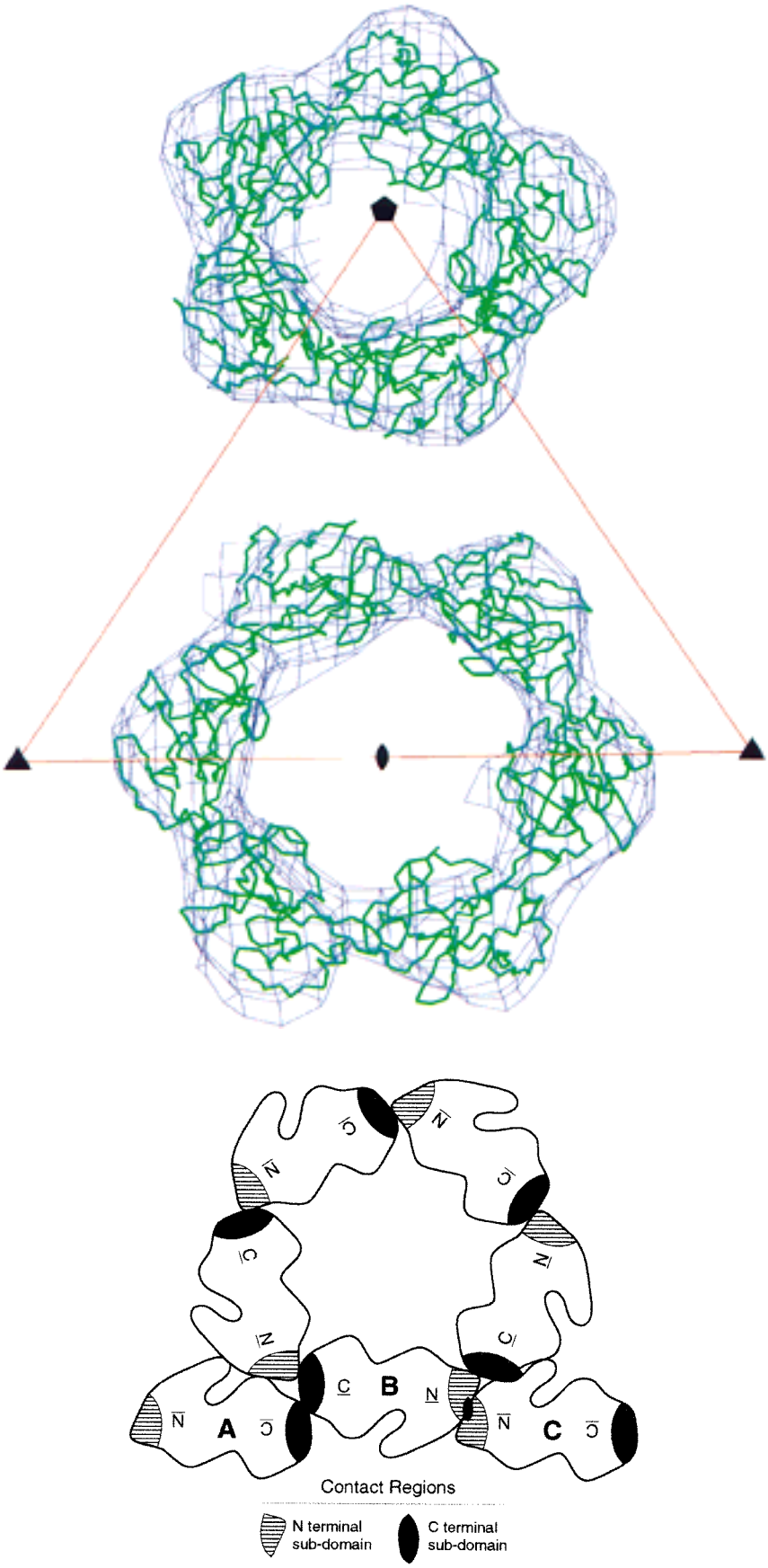


Fig. 3. Monomer contacts in the AB and BC crystallographic dimers, as well as in the virion. **a:** Fit of SFCP into the cryo-EM density of Ross River virus. The pentamers and hexamers are in a $T = 4$ surface lattice. Icosahedral symmetry axes are shown. **b:** Association of SFCP subunits within a pentamer of the virion and in the ABC trimer found in crystal packing.

TABLE V. Monomer-Monomer Interactions

A. Monomer—Monomer Contacts in SFCP Crystal Structures

Crystal type	Contact	Residues*
I	AB	**189–194, 198–199, 226**
I	BC	**147–149, 157–162, 164**
II	AB	**189–194, 198–199, 226**
II	BA′ (~BC)†	**157–162, 164**

*Each residue implicated in the contact has at least one atom less than 3.8 Å from an atom in the neighboring subunit.
†In type II crystals, there is a continuous chain of monomers ABA′ . . . in the crystal lattice. The BA′ contacts are similar to the BC contacts in crystal form I.

B. Monomer—Monomer Contacts in the Virion Head, based on EM Image Reconstruction and Fitting of X-ray Structure

Contact	Residues* in amino-terminal sub-domain	Residues[a] in carboxy-terminal sub-domain
Pentamer	**122–132, 147–152, 156, 215–217**	**185–195, 198–202, 207, 222–232**
Hexamer	**122–130, 149, 151, 215**	**185–194, 200, 202, 207, 215, 223–229**

*Each residue implicated in the contact has at least one atom less than 3.8 Å from an atom in the neighboring subunit.

E_1, F_1 and A_2, B_2, C_2, D_2, E_2, F_2) in the characteristic fold of chymotrypsin-like serine proteinases. The stability of the structure is enhanced by the conserved aromatic residues **Tyr193, Trp195, Tyr202,** and **Phe207** buried within the carboxy-terminal β barrel. The overall structure is considerably more compact than chymotrypsin or the other serine proteinases, with some β strands truncated and many of the connecting loops drastically shortened, but the topology of the β barrels is retained. Similar to SCP,[8] SFCP has two extra β strands, G_2 and H_2, at the C-terminus of the protein, taking the place of the α helix found in the mammalian serine proteinases. This difference changes the path of the polypeptide chain when compared to chymotrypsin, so that the C-terminus is near the center of the molecule and is in the active center. The catalytic triad, consisting of **Ser219, His145,** and **Asp167,** is found between the two β barrel subdomains. The C-terminal residue, **Trp267,** lies in the large hydrophobic pocket corresponding to the specificity pocket of chymotrypsin, inhibiting further proteolytic activity after cleavage of the peptide bond between **Trp267** and **Ser 268.**

There is another hydrophobic pocket adjacent to the active center, also located between the two subdomains, composed primarily of **Met141** on βC_1, **Cys170** on βF_1, **Tyr184** on the loop connecting the two subdomains and **Trp251** on βF_2. This pocket may be associated with the binding of the carboxy-terminal cytoplasmic domain of the E2 glycoprotein.[24,25]

Crystal Content

A recurring concern in the study of full-length alphavirus core proteins has been to determine whether proteolytic cleavage occurred during the many months required for the crystallization procedure. It is, therefore, possible that the amino-terminal domain is not disordered, but has been cleaved prior to crystal formation. The V_M values (Table II) are at the lower limit of what is normally observed in other protein crystals.[27] The above concern was reinforced by the observation that SCP (residues *106–264*) expressed in *E. coli*[37] crystallizes isomorphously with type 2 crystals characterized for full-length SCP, suggesting that the full-length wild-type (WT) protein has been cleaved to a size similar to the recombinant truncated protein.

It was reported earlier[38] that the amino end of the protein from the crystallization mixture of SCP was found to be blocked. That would be consistent with the presence of full-length WT protein. However, molecular weight determination of SFCP from dissolved crystals using mass spectroscopy clearly established that the protein had been cleaved between residues **108** and **109.**

Comparison of Monomeric Structures

Although noncrystallographic symmetry restraints were used in the refinement of the SFCP crystal structures, monomers A and B were found to be more alike to each other than they were to monomer C. The rms deviation between equivalent C_α positions between monomers A and B was 0.18 Å in type I crystals and 0.19 Å in type II crystals, while the rms difference was 0.33 Å between monomers A and C or 0.31 Å between monomers B and C in type I crystals. There are three regions of biggest flexibility, two of which occur where SFCP has a one-residue deletion relative to SCP.

The region between residues **252** and **255** differs by as much as 3 Å for the C_α positions in monomer C

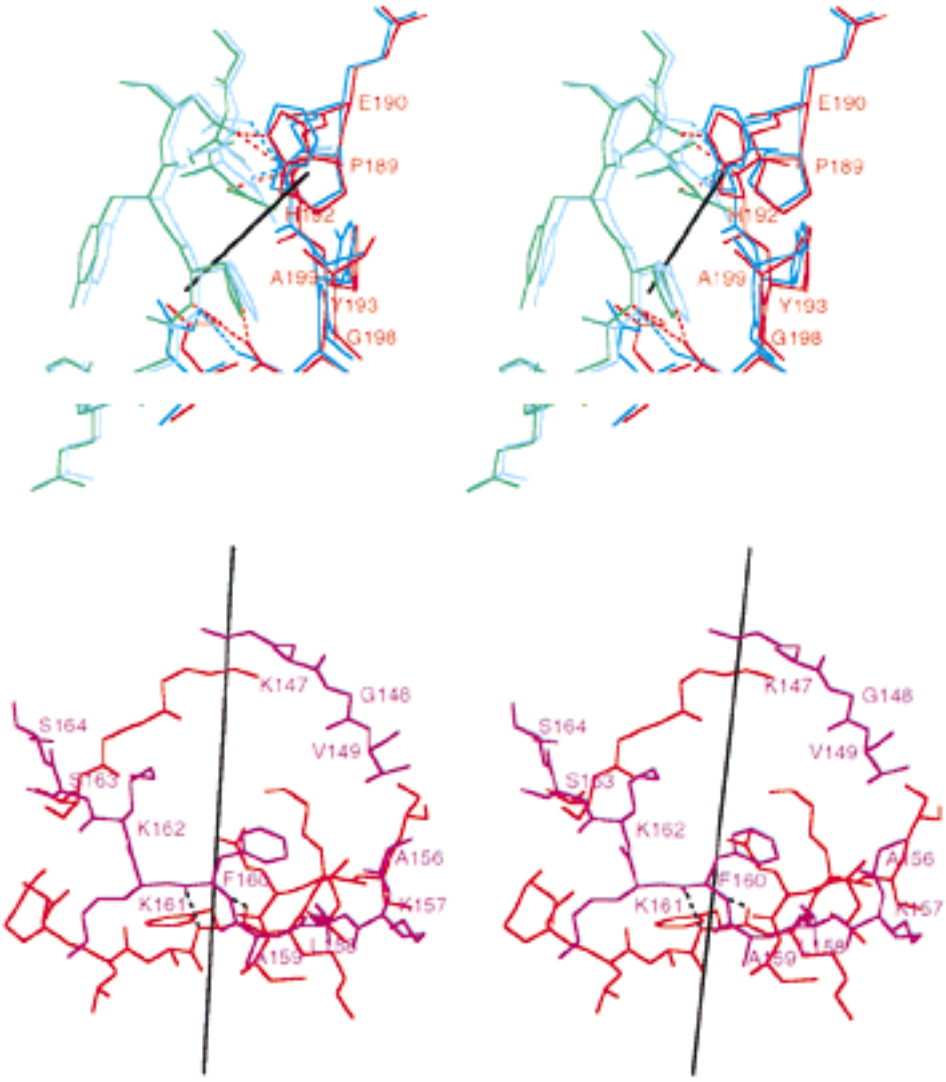


Fig. 4. Dimer interfaces. Twofold axes are shown as black lines. The SFCP monomers A, B, and C are shown in the same color convention as in Figure 2. **a:** The AB dimer interface showing the SCP structure (type 2 crystals) in sky blue and dark blue superimposed onto SFCP (type I crystals) in green and red, respectively. **b:** The BC dimer interface in SFCP crystal type I in red and purple.

relative to either monomer A or B in type I crystals. Comparison between monomers A and B in type I crystals with those in type II crystals shows differences of as much as 1.6 Å between the position of equivalent C_α atoms in this same region. Similarly, in comparing the SCP structure in type 2 crystals with other SCP structures and with SFCP structures, the largest differences occur in this region (Fig. 5). These residues are involved in lattice contacts between dimers and are also associated with the hydrophobic pocket, where E2 is proposed to bind.

Another large difference between the A or B monomers of SFCP type I or II crystals and SCP monomers occurs between residues **His125** and **Lys128,** where SFCP has a one-residue deletion between the A_1 and B_1 β strands (Fig. 1). These residues form a loop on the surface of the protein and are located away from the specificity pocket. This loop is involved in forming the head-to-tail contacts in hexamer and pentamer formation in virions[4] (Fig. 3).

The third region of flexibility (residues **171–181**) occurs in the connecting region between the two subdomains and also involves a part of the hydrophobic pocket for E2 binding. This region is particularly variable among the SCP structures.

When the structure of SFCP is superimposed on SCP fitted into the cryo-EM density of the homologous Ross River virus (RRV),[4] the SCP inserted

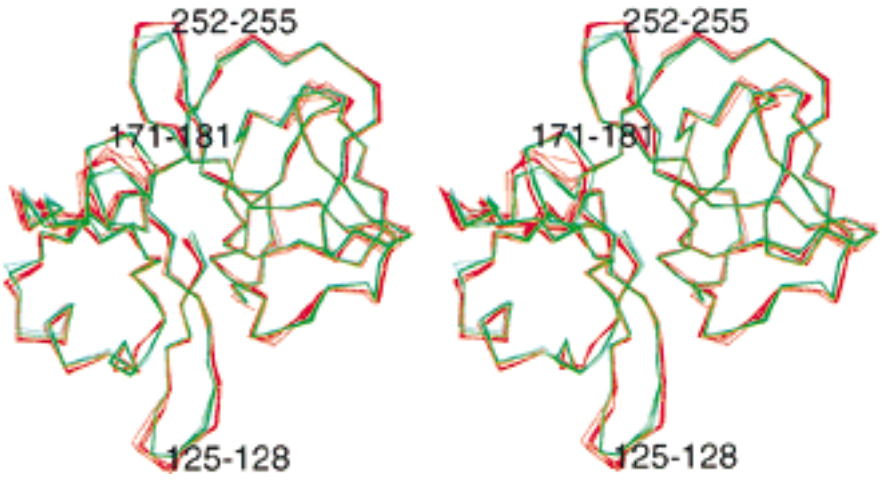


Fig. 5. C_α drawing comparing five different SCP structures (red) distributed among the types 2, 3, and 4 crystal forms with five different SFCP structures (green) distributed among the types I and II crystals. The three regions of greater flexibility are labeled with SFCP numbers.

residue *250* is seen to be outside the density. The other insertion site in SCP at residue *121* causes steric hindrance between adjacent monomers. In contrast, the SFCP structure is well within the density and there is little steric clash between neighboring subunits. The original fit was established with the SCP structure, as the SFCP structure had not yet been determined. The sequence of RRV core protein shows greater similarity to SFCP than it does to SCP, including the site of deletions relative to SCP. This improvement of the fit further confirms the accuracy of the core protein structure placement within the cryo-EM density.

The structural comparisons show that the functional surfaces of the capsid protein (capsid protein-capsid protein interactions and capsid protein-glycoprotein E2 spike interactions) are the most variable in structure. This is in contrast to the active centers in enzymes, which are usually the most conserved part of the structure in amino acid composition and conformation.

The Putative E2 Binding Site

The binding site of the C-terminal sequence of the E2 glycoprotein has been proposed to be a hydrophobic pocket[24,25] on the surface of the alphavirus capsid protein. In type 2, 3, and 4 SCP structures, this pocket binds the capsid sequence *108–111* (the "amino-terminal arm") of neighboring molecules in a manner that is probably analogous to the binding of E2 to cores.[25] However, the equivalent residues **112–116** of SFCP are not visible in the electron density. The crystal packing arrangement does not permit the amino-terminal arm to reach into the pockets of the neighboring molecules or into the pocket of its own monomer. Hence, the amino-terminal arm does not bind into the hydrophobic pocket in either crystal structure I or II (Fig. 6). The major difference in the pocket conformation in SFCP compared to SCP is that **Tyr184** is exposed in SFCP, whereas in SCP, with the bound amino-terminal arm in the pocket, this conserved tyrosine is buried (Fig. 6). Coombs and Brown[39] have shown that only this tyrosine (*Tyr180*) can be iodinated in SINV cores, indicating that the tyrosine is external in the absence of bound glycoproteins. Thus, the hydrophobic pocket in SFCP may have the conformation of the unbound protein, suggesting that binding of the glycoprotein E2 to the assembled nucleocapsid core induces a large conformational change. Exposure of **Tyr184** in SFCP creates an empty space in the pocket, which is at least partially filled by a shift in residue **Trp251.** A similar conformational change of this tryptophan was also observed for an SCP mutant, in which the conserved tyrosine was replaced by a smaller serine residue.[25,37]

The Catalytic Center

SFCP has a similar arrangement of the three essential catalytic residues, **His145, Asp167,** and **Ser219,** to that observed in SCP. The C-terminal **Trp267** lies in the specificity pocket, as in SCP (Fig. 7). The residues which comprise the specificity pocket are mostly conserved in all known alphavirus sequences, except for residues **209, 213,** and **215.** Of the two carboxy oxygen atoms at the C-terminus, one is hydrogen-bonded to $N_{\epsilon 2}$ of **His145,** while the other is close (3.1 to 3.7 Å) to the oxyanion hole formed by the main chain amino groups of residues **Gly217** and **Ser219.**

Surface of the Viral Cores

Cryo-EM reconstruction of RRV,[4] an alphavirus closely related to SFV, shows that the protein cores have a $T = 4$ surface lattice consisting of hexameric and pentameric units (Fig. 3). Previous results have

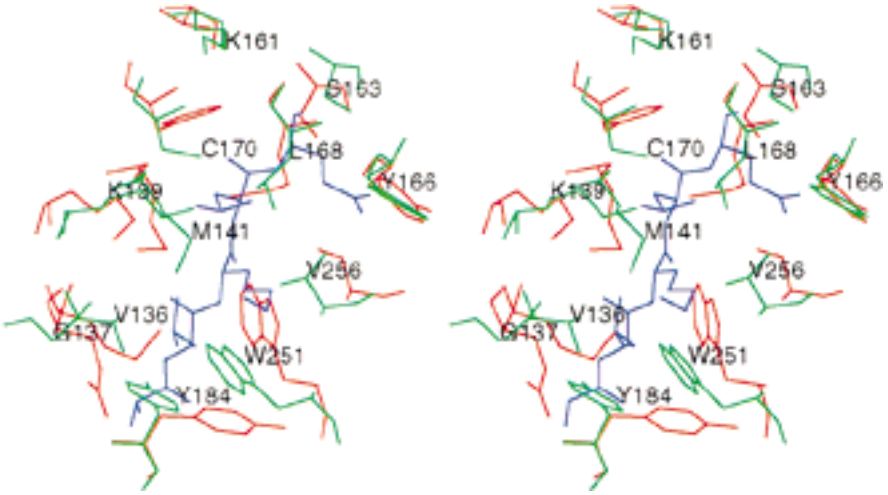


Fig. 6. Comparison of the hydrophobic pocket in SFCP (green) (monomer A in type I crystals) and SCP (red) (type 2 crystals). Note that **Tyr184** in SFCP is pointing out of the pocket. The SCP amino-terminal residues *107–111* are shown in blue. The SFCP pocket is empty.

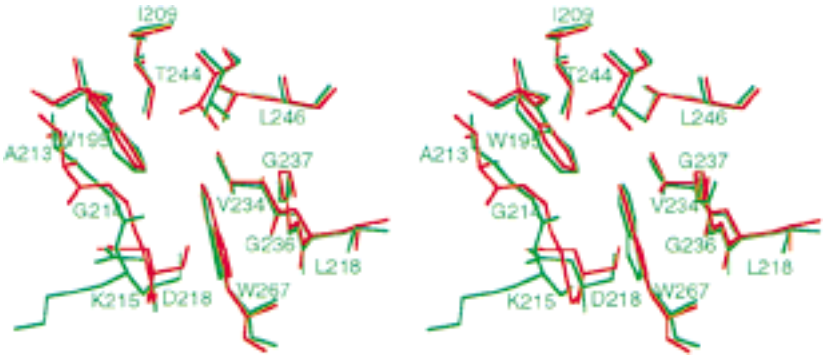


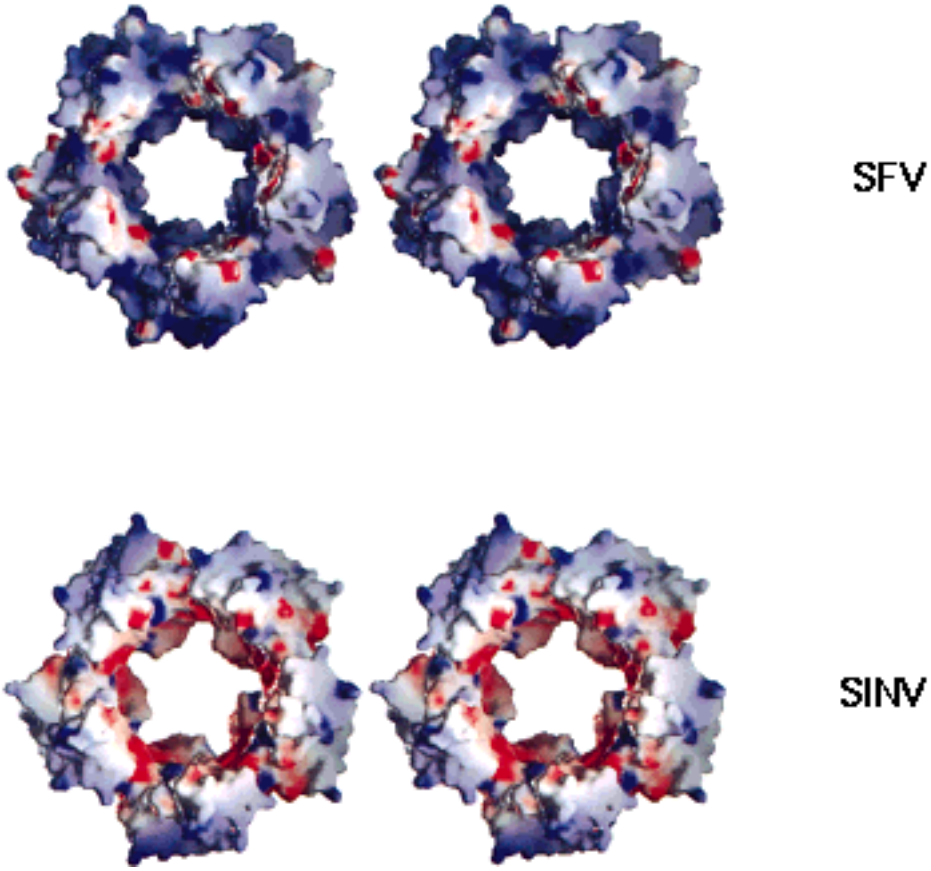
Fig. 7. The specificity pocket showing that the SCP structure (green) (monomer B in type I crystals) and SCP structure (red) (type 2 crystals), derived from cleaved WT protein, are similar.

established the approximate fit of the capsid structure into the cryo-EM density.[4] This permits a comparison of the surface charge distributions of SFV and SINV (Fig. 8a). The area exposed by each surface amino acid is depicted as a "roadmap" (Fig. 8b).[40,41] The central ring of residues on the capsid exterior, which must be buried in the lipid bilayer, is primarily hydrophobic. These residues have a higher than average level of conservation. Surrounding the hydrophobic residues are positively-charged residues, which could interact with the negatively charged phosphate head groups within the lipid. The amount of positive charge is considerably greater for SFV than for SINV. The internal surface has more negatively-charged residues than the external surface, particularly for SINV. As the internal surface should be in contact with the viral RNA, it is possible that the positively charged, disordered, amino-terminal domain of the capsid protein interfaces between the RNA and the protein shell.

CONCLUSIONS

The organization of the capsid protein in alphaviruses had been established by cryo-EM image reconstruction[3,4] as a $T = 4$ surface lattice consisting of pentamers and hexamers. Surprisingly, however, most crystal forms of SCP and SFCP contain the same "AB" dimer (with tail-to-tail contacts between the carboxy-terminal subdomains), which does not exist in the virion.[4] The structure of SFCP now shows a second type of "BC" dimer (with head-to-head contacts between the amino-terminal subdomains) present in both structure determinations reported here. In the virion, there are tail-to-head contacts using the surfaces found in the carboxy-terminal subdomain contact region of AB dimers and the amino-terminal subdomain contact region of BC dimers.

The structure of SFCP, like the structure of SCP, has a hydrophobic surface pocket, which has been



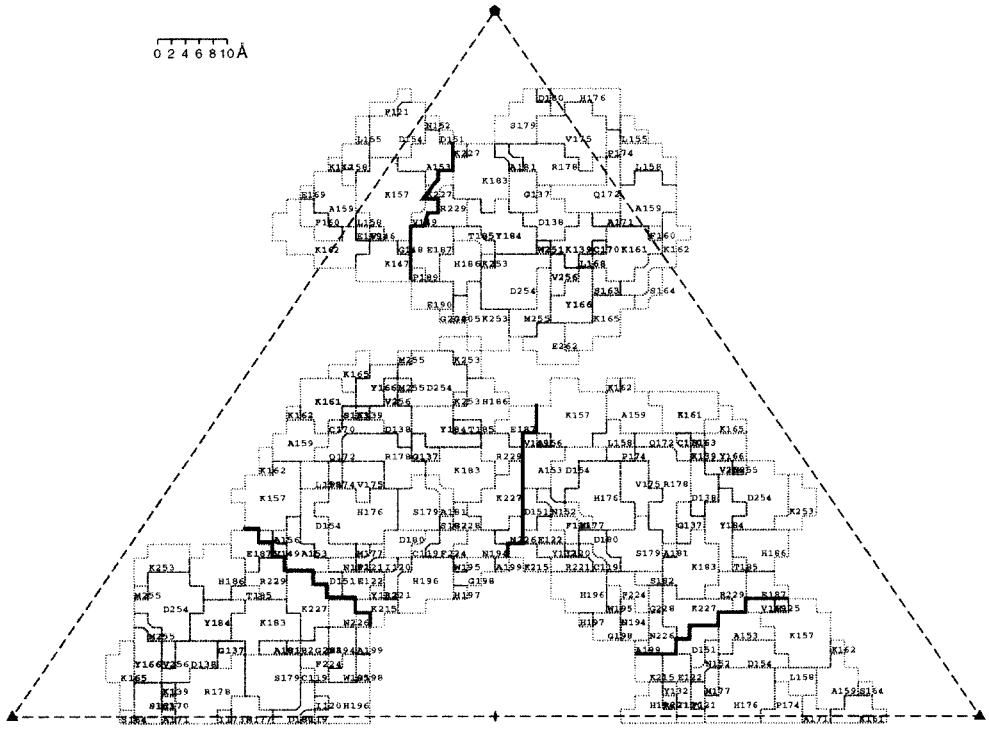


Fig. 8. **top:** Comparison of the charge distribution for the outer surface of the core in the SFV (top) and SINV (bottom) pentamer units. The relation of color to potential ranges from −10 (red) to +10 (purple) kT. The calculations were produced by the GRASP program.[43] **bottom:** The detailed amino acid distribution for SFV on the external surface of the core can be seen in a ''road-map.''[40,41] The demarcation between monomers is indicated with bold lines. The residues forming a surface hydrophobic pocket, the putative E2 binding site, are shaded.

implicated in binding the carboxy end of the E2 glycoprotein.[24,25] However, the pocket in the SFCP structures is empty, whereas in SCP structures the pocket binds the capsid residues *108* to *111* from a neighboring molecule. These capsid residues are similar to the residues in the E2 cytoplasmic domain which are known to be required for viral budding. In the empty SFCP pocket, the **Tyr184** is pointing out, while in the SCP pocket with the bound peptide, the homologous *Tyr180* is pointing into the pocket. This suggests that binding of E2 to the capsid protein during viral budding produces a conformational change involving the burial of **Tyr184,** consistent with iodination data of this tyrosine.[39] Burial of the tyrosine on viral maturation or exposure of the tyrosine on viral cell entry may be analogous to the regulation of viral stability of rhinoviruses and polioviruses by a hydrophobic "pocket factor."[42]

ACKNOWLEDGMENTS

We are grateful for helpful discussions with Richard Kuhn and Katey E. Owen, and thank Cheryl Towell and Sharon Wilder for help in the preparation of this manuscript. The work was supported by grants from the National Science Foundation (MCB-9102855) to M.G.R., by the Deutsch Forschungsgemeinschaft (SFB 272) to G.W. and, in part, a National Institutes of Health program project grant (AI 35312) to M.G.R.

REFERENCES

1. Strauss, J.H., Strauss, E.G. The alphaviruses: Gene expression, replication, and evolution. Microbiol. Rev. 58:491–562, 1994.
2. Shope, R.E. Medical significance of togaviruses: An overview of diseases caused by togaviruses in man and in domestic and wild vertebrate animals. In: "The Togaviruses. Biology, Structure, Replication." Schlesinger, R.W. (ed.) New York: Academic Press, 1980:47–82.
3. Fuller, S.D., Berriman, J.A., Butcher, S.J., Gowen, B.E. Low pH induces swiveling of the glycoprotein heterodimers in the Semliki Forest virus spike complex. Cell 81:715–725, 1995.
4. Cheng, R.H., Kuhn, R.J., Olson, N.H., Rossmann, M.G., Choi, H.K., Smith, T.J., Baker, T.S. Nucleocapsid and glycoprotein organization in an enveloped virus. Cell 80: 621–630, 1995.
5. Strauss, E.G., Strauss, J.H. Structure and replication of the Alphavirus genome. In: "The Togaviridae and Flaviviridae." Schlesinger, S., Schlesinger, M.J. (eds.) New York: Plenum, 1986:35–90.
6. Hahn, C.S., Strauss, E.G., Strauss, J.H. Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid protein autoprotease. Proc. Natl. Acad. Sci. USA 82:4648–4652, 1985.
7. Melancon, P., Garoff, H. Processing of the Semliki Forest virus structural polyprotein: Role of the capsid protease. J. Virol. 61:1301–1309, 1987.
8. Choi, H.K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M.G., Wengler, G. Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. Nature 354:37–43, 1991.
9. Tong, L., Choi, H.K., Minor, W., Rossmann, M.G. The structure determination of Sindbis virus core protein using isomorphous replacement and molecular replacement averaging between two crystal forms. Acta Crystallogr. A48:430–442, 1992.
10. Hahn, C.S., Strauss, J.H. Site-directed mutagenesis of the proposed catalytic amino acids of the Sindbis virus capsid protein autoprotease. J. Virol. 64:3069–3073, 1990.
11. Tong, L., Wengler, G., Rossmann, M.G. Refined structure of Sindbis virus core protein and comparison with other chymotrypsin-like serine proteinase structures. J. Mol. Biol. 230:228–247, 1993.
12. Shirako, Y., Niklasson, B., Dalrymple, J.M., Strauss, E.G., Strauss, J.H. Structure of the Ockelbo virus genome and its relationship to other Sindbis viruses. Virology 182:753–764, 1991.
13. Levinson, R.S., Strauss, J.H., Strauss, E.G. Complete sequence of the genomic RNA of O'Nyong-nyong virus and its use in the construction of alphavirus phylogenetic trees. Virology 175:110–123, 1990.
14. Kinney, R.M., Johnson, B.J.B., Brown, V.L., Trent, D.W. Nucleotide sequence of the 26S mRNA of the virulent Trinidad donkey strain of Venezuelan equine encephalitis virus and deduced sequence of the encoded structural proteins. Virology 152:400–413, 1986.
15. Hahn, C.S., Lustig, S., Strauss, E.G., Strauss, J.H. Western equine encephalitis virus is a recombinant virus. Proc. Natl. Acad. Sci. USA 85:5997–6001, 1988.
16. Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H., Delius, H. The capsid protein of Semliki Forest virus has clusters of basic amino acids and proteins in its amino-terminal region. Proc. Natl. Acad. Sci. USA 77:6376–6380, 1980.
17. Faragher, S.G., Meek, A.D.J., Rice, C.M., Dalgarno, L. Genome sequences of a mouse-avirulent and a mouse-virulent strain of Ross River virus. Virology 163:509–526, 1988.
18. Chang, G.-J., Trent, D.W. Nucleotide sequence of the genome region encoding the 26S mRNA of eastern equine encephalomyelitis virus and the decoded amino acid sequence of the viral structural proteins. J. Gen. Virol. 68:2129–2142, 1987.
19. Boege, U., Wengler, G., Wengler, G., Wittmann-Liebold, B. Primary structure of the core proteins of the alphaviruses Semliki Forest virus and Sindbis virus. Virology 113:293–303, 1981.
20. Zhao, H., Lindqvist, B., Garoff, H., von Bonsdorff, C., Liljeström, P. A tyrosine-based motif in the cytoplasmic domain of the alphavirus envelope protein is essential for budding. EMBO J. 13:4204–4211, 1994.
21. Lee, H., Brown, D.T. Mutations in an exposed domain of Sindbis virus capsid protein result in the production of noninfectious virions and morphological variants. Virology 202:390–400, 1994.
22. Metsikkö, K., Garoff, H. Oligomers of the cytoplasmic domain of the p62/E2 membrane protein of Semliki Forest virus bind to the nucleocapsid in vitro. J. Virol. 64:4678–4683, 1990.
23. Vaux, D.J.T., Helenius, A., Mellman, I. Spike-nucleocapsid interaction in Semliki Forest virus reconstructed using network antibodies. Nature 336:36–42, 1988.
24. Skoging, U., Vihinen, M., Nilsson, L., Liljeström, P. Aromatic interactions define the binding of the alphavirus spike to its nucleocapsid. Structure 4:519–529, 1996.
25. Lee, S., Owen, K.E., Choi, H.K., Lee, H., Lu, G., Wengler, G., Brown, D.T., Rossmann, M.G., Kuhn, R.J. Identification of a protein binding site on the surface of the alpha-virus nucleocapsid and its implication in virus assembly. Structure 4:531–541, 1996.
26. Kabsch, W. Evaluation of single-crystal x-ray diffraction data from a position-sensitive detector. J. Appl. Crystallogr. 21:916–924, 1988.
27. Matthews, B.W. Solvent content of protein crystals. J. Mol. Biol. 33:491–497, 1968.
28. Ulmanen, I., Söderlund, H., Kääriäinen, L. Semliki Forest virus capsid protein associates with the 60S ribosomal subunit in infected cells. J. Virol. 20:203–210, 1976.
29. Rossmann, M.G., Blow, D.M. The detection of sub-units within the crystallographic asymmetric unit. Acta Crystallogr. 15:24–31, 1962.

30. Tong, L., Rossmann, M.G. The locked rotation function. Acta Crystallogr. A46:783–792, 1990.
31. Tong, L., Rossmann, M.G. Patterson-map interpretation with noncrystallographic symmetry. J. Appl. Crystallogr. 26:15–21, 1993.
32. Rossmann, M.G. The molecular replacement method. Acta Crystallogr. A46:73–82, 1990.
33. Brünger, A.T., Kuriyan, J., Karplus, M. Crystallographic *R* factor refinement by molecular dynamics. Science 235:458–460, 1987.
34. Navaza, J. AMoRe: an automated package for molecular replacement. Acta Crystallogr. A50:157–163, 1994.
35. Engh, R.A., Huber, R. Accurate bond and angle parameters for x-ray protein structure refinement. Acta Crystallogr. A47:392–400, 1991.
36. Crowther, R.A. The fast rotation function. In: "The Molecular Replacement Method." Rossmann, M.G. (ed.) New York: Gordon & Breach, 1972:173–178.
37. Choi, H.K., Lee, S., Zhang, Y.P., McKinney, B.R., Wengler, G., Rossmann, M.G., Kuhn, R.J. Structural analysis of Sindbis virus capsid mutants involving assembly and catalysis. J. Mol. Biol. 262:151–167, 1996.
38. Boege, U., Cygler, M., Wengler, G., Dumas, P., Tsao, J., Luo, M., Smith, T.J., Rossmann, M.G. Sindbis virus core protein crystals. J. Mol. Biol. 208:79–82, 1989.
39. Coombs, K., Brown, D.T. Topological organization of Sindbis virus capsid protein in isolated nucleocapsids. Virus Res. 7:131–149, 1987.
40. Chapman, M.S. Mapping the surface properties of macromolecules. Prot. Sci. 2:459–469, 1993.
41. Rossmann, M.G., Palmenberg, A.C. Conservation of the putative receptor attachment site in picornaviruses. Virology 164:373–382, 1988.
42. Rossmann, M.G. Viral cell recognition and entry. Prot. Sci. 3:1712–1725, 1994.
43. Nicholls, A., Sharp, K.A., Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 11:281–296, 1991.