# Identification of a protein binding site on the surface of the alphavirus nucleocapsid and its implication in virus assembly

Sukyeong Lee[1], Katherine E Owen[1], Hok-Kin Choi[1], Heuiran Lee[2], Guoguang Lu[1], Gerd Wengler[3], Dennis T Brown[2], Michael G Rossmann* and Richard J Kuhn*

**Background**: Many enveloped viruses exit cells by budding from the plasma membrane. The driving force for budding is the interaction of an inner protein nucleocapsid core with transmembrane glycoprotein spikes. The molecular details of this process are ill defined. Alphaviruses, such as Sindbis virus (SINV) and Semliki Forest virus (SFV), represent some of the simplest enveloped viruses and have been well characterized by structural, genetic and biochemical techniques. Although a high-resolution structure of an alphavirus has not yet been attained, cryo-electron microscopy (cryo-EM) has been used to show the multilayer organization at 25 Å resolution. In addition, atomic resolution studies are available of the C-terminal domain of the nucleocapsid protein and this has been modeled into the cryo-EM density.

**Results**: A recombinant form of Sindbis virus core protein (SCP) was crystallized and found to diffract much better than protein extracted from the virus (2.0 Å versus 3.0 Å resolution). The new structure showed that amino acids 108 to 111 bind to a specific hydrophobic pocket in neighboring molecules. Re-examination of the structures derived from virus-extracted protein also showed this 'N-terminal arm' binding to the same hydrophobic pocket in adjacent molecules. It is proposed that the binding of these capsid residues into the hydrophobic pocket of SCP mimics the binding of E2 (one of two glycoproteins that penetrate the lipid bilayer of the viral envelope) C-terminal residues in the pocket. Mutational studies of capsid residues 108 and 110 confirm their role in capsid assembly.

**Conclusions**: Structural and mutational analyses of residues within the hydrophobic pocket suggest that budding results in a switch between two conformations of the capsid hydrophobic pocket. This is the first description of a viral budding mechanism in molecular detail.

Addresses: [1]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA, [2]The Cell Research Institute and Department of Microbiology, The University of Texas, Austin, TX 78713-7640, USA and [3]Institut für Virologie, Justus-Liebig-Universität Giessen, Frankfurter Strasse 108, D-6300 Giessen, Germany.

*Corresponding authors.

Key words: assembly, capsid structure, mutational analysis, Sindbis virus, virus budding

Received: 29 Jan 1996
Revisions requested: 20 Feb 1996
Revisions received: 28 Feb 1996
Accepted: 29 Feb 1996

Structure 15 May 1996, 4:531–541

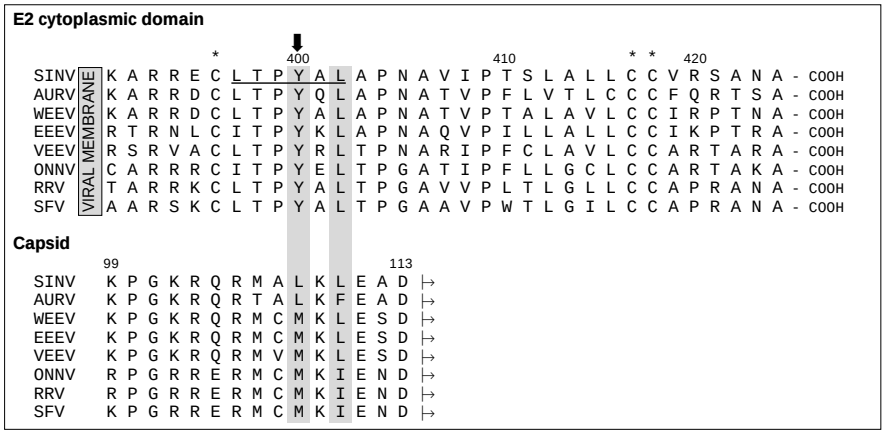
 ISSN 0969-2126

## Introduction

Alphaviruses are a group of mosquito-borne, enveloped viruses which can cause encephalitis, fever, arthritis and rash in mammals. The nucleocapsid core has an external radius of about 205 Å. It consists of a positive-strand RNA genome surrounded by 240 copies of the capsid protein arranged in a $T$=4 surface lattice [1–3]. The core is enveloped by a lipid bilayer that is penetrated by 80 glycoprotein spikes. These spikes are also arranged in a $T$=4 lattice that is in register with the internal nucleocapsid. Each spike is a trimer of heterodimers that extends to a viral radius of 345 Å with a receptor attachment site on its globular extremity [4]. The heterodimer consists of two glycoproteins, E1 and E2, whose C-terminal ends protrude on the cytoplasmic side of the membrane by 2 amino acids and between 31 and 33 amino acids, respectively.

The structural proteins (capsid/E3/E2/6K/E1) are translated as a polyprotein from a subgenomic RNA. The polyprotein is post-translationally cleaved into component proteins by viral and cellular proteases. Following self-cleavage from the nascent polyprotein, the capsid protein specifically binds the replicated genomic RNA, and assembly of the nucleocapsid occurs in the cytoplasm soon after capsid protein synthesis [5]. The spike proteins are glycosylated in the Golgi and are then associated with the nucleocapsid at the plasma membrane during budding [6,7]. The translocation of the glycoproteins across the membrane is regulated by various signal sequences and further post-translational modifications, including palmitoylation of some or all of the cysteine residues, 396, 416 and 417 (all amino acid sequence numbers will refer to Sindbis virus [SINV], unless otherwise stated), in the cytoplasmic C terminus of E2 [8–10].

**Figure 1**




Alignment of the E2 cytoplasmic domain with the capsid protein. Numbering corresponds to SINV. Shaded residues in both proteins are important in binding to the pocket. Underlined residues represent the minimal peptide sequence found to inhibit viral budding ([19]; see text ). Residues that are palmitoylated are indicated by a *; → represents the C-terminal catalytic domain. The sequences were obtained from Strauss and Strauss [5], Rümenapf *et al.* [41], and Wengler *et al.* [31]: SINV, Sindbis virus; AURV, Aura virus; WEEV, Western equine encephalitis virus; EEEV, Eastern equine encephalitis virus; VEEV, Venezuelan equine encephalitis virus; ONNV, O'nyong-nyong virus; RRV, Ross River virus; SFV, Semliki Forest virus.

The interaction between the nucleocapsid and the C-terminal E2 [11,12], but not E1, residues [13], is essential for viral budding [14–16]. Site-specific mutagenesis of the conserved Tyr400 in the E2 glycoprotein C-terminal (cytoplasmic) domain of SINV and Semliki Forest virus (SFV) identified the approximate site of interaction with the nucleocapsid (Fig. 1) [9,17]. Chimeric viruses of SINV and Ross River virus (RRV) showed that Ala403 and Asn405 in the E2 cytoplasmic domain are also important for budding [18]. Collier *et al.* [19] showed that peptides corresponding to sequences of the E2 cytoplasmic domain of SINV (Fig. 1) could inhibit virion formation. They postulated that these peptides competitively inhibited attachment of E2 to the nucleocapsid. These peptides include sites subsequently identified by mutagenesis as important for budding [9,17]. Site-specific mutagenesis of the capsid protein itself identified Tyr180 and/or Glu183 as probable participants in the binding of the nucleocapsid to E2 [20,21].

The structures of SINV capsid protein (SCP) [22,23] and SFV capsid protein, SFCP, (H-KC, GL, SL, GW and MGR, unpublished data) have been determined to at least 3 Å resolution. Their C-terminal domains, Arg114 to Trp264, were found to have a fold similar to chymotrypsin-like serine proteinases. SCP contains the catalytic triad Ser215, His141 and Asp163. Consistent with earlier observations [24–27], the structure of SCP showed that this protein cleaved itself from its own polyprotein, leaving the C-terminal tryptophan in the substrate-binding pocket, thereby inhibiting further catalytic activity. The first 113 residues of SCP were suggested to be disordered in the crystal structure. Many of these residues were basic and were presumed to be associated with the viral RNA in the virion as is common in many plant viruses [28]. Some of these residues have been shown to be involved in the specific recognition of RNA [29–31].

The most detailed structure determinations of whole alphaviruses were reported by Cheng *et al.* [1], for RRV, and by Fuller *et al.* [2], for SFV, using cryo-EM at about 25 Å resolution. Combination of the X-ray crystallographic results for SCP and the cryo-EM results for RRV gave a convincing fit of the ordered, C-terminal domain structure of the capsid protein into the pentameric and hexameric units of the nucleocapsid core found in the whole virus [1].

The chymotrypsin-like structure of the C-terminal domain of SCP and SFCP consists of two subdomains, with each subdomain having a 6- or 7-stranded, antiparallel, β-barrel ('Greek key') structure. The active site is in a cleft between these two sub-domains. Previous crystallographic studies found dimers for both SCP and SFCP in which the C-terminal subdomains of each monomer were in tail-to-tail contact [22]. In one structure of SFCP (H-KC, GL, SL, GW and MGR, unpublished data), there is also a head-to-head dimer in which the N-terminal subdomains are in contact (Table 1). Neither kind of dimer occurs in the native virus, but rather there are head-to-tail contacts that make pentamers and hexamers in the nucleocapsid cores [1]. The fit of the C-terminal domain of SCP into the RRV cryo-EM density shows that the head-to-tail contacts in the virion are the result of binding the head-to-head contact region with the tail-to-tail contact regions seen in the crystallographic studies.

We have now determined a structure of recombinant SCP (residues 106–266 with Ser215→Ala). The protein was originally investigated to determine the position of the P1′ and P2′ sites prior to proteolytic cleavage. This structure showed that the N-terminal residues 108–111 bind into a hydrophobic pocket on the surface of a neighboring molecule. Re-examination of the earlier, lower resolution results for SCP showed that in every case the pocket was also occupied by the same 'N-terminal arm'. However, in

**Table 1**

**Alphavirus capsid protein crystal forms.**

| | Crystal form* | Residues | Mutant form | Number of monomers per crystallographic asymmetric unit | Nature of dimerization | Donor monomer† | Symmetry operation | Acceptor monomer |
|---|---|---|---|---|---|---|---|---|
| SCP‡ | Type 2 | 1–264 | wild type | | | | | |
| | | 106–264 | wild type | | | | | |
| | | 106–264 | Ser215→Ala | 1 | tail to tail | A | ½+y, 1½–x, ¼+z | A |
| | | 106–264 | Tyr180→Ser; Glu183→Gly | | | B | 1½–y, ½+x, –¼+z | B |
| SCP | Type 3 | 1–264 | wild type | 2 | tail to tail | B | 1–x, –½+y, 1–z | A |
| | | | | | | A | 1–x, ½+y, –z | B |
| SCP | Type 4 | 106–266 | Ser215→Ala | 2 | head to tail | A | 1+x, y, z | A |
| | | | | | | B | 1+x, y, z | B |
| SFCP# | Type I | 1–267 | | 3 | tail-to-tail (A,B) and head-to-head (B,C) | | – | – |
| SFCP | Type II | 1–267 | | 2 | tail-to-tail | | – | – |

*See text for definition of crystal forms. †The donor molecule is defined as having its N-terminal arm (residues 108–111) bound to the hydrophobic pocket of the acceptor molecule. Application of the given symmetry operation will superimpose the acceptor molecule onto the donor molecule.‡Sindbis virus core protein; #Semliki forest virus capsid protein.

two SFCP crystal structures the pocket is unoccupied (H-KC, GL, SL, GW and MGR, unpublished data). In addition, there is a significant difference between the conformations of the pocket in SCP and in SFCP, although the important residues involved in forming this pocket are identical for both viruses. Mutagenesis of residues 108 and 110 in SCP (this work) produces viruses that are deficient in nucleocapsid core accumulation and suggests that the binding seen in the crystallographic studies may have a biological role in assembly. Furthermore, we show that this binding mode is probably analogous to the way in which the E2 C-terminal cytoplasmic residues bind to the nucleocapsid core. The position of the binding pocket on the nucleocapsid, determined by cryo-EM results and X-ray crystallography [1], is consistent with the site of penetration of the glycoprotein spike through the plasma membrane being close to the pocket. We propose that a switch between the two conformations of the hydrophobic pocket occurs on binding E2 and the energy derived from the burial of the capsid protein and E2 hydrophobic surfaces drives the budding process of the mature virion.

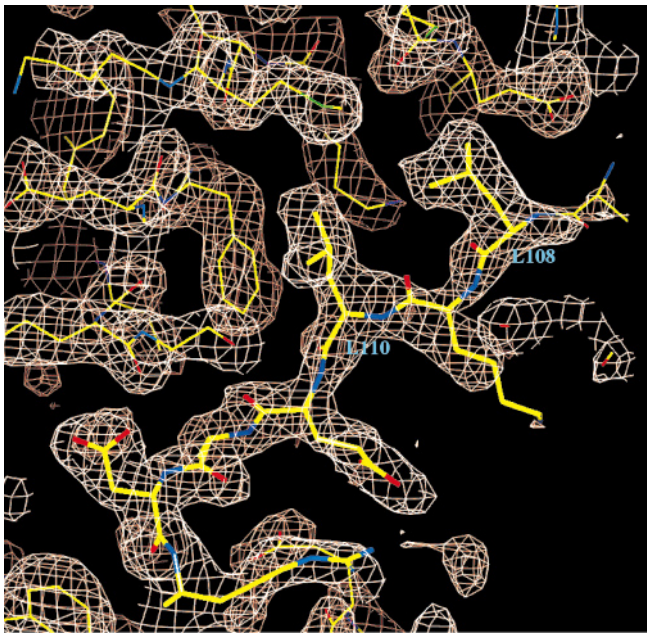
## Results and discussion

### Binding of N-terminal arm residues 108–111 to SCP

A mutated form of SCP was expressed in *E. coli* (H-KC, *et al.*, & RJK, unpublished data). The mutant protein had the essential Ser215 altered to alanine, resulting in inhibition of catalytic activity. In addition, residues Ser265 and Ala266, present in the polyprotein, were retained in order to study the substrate when bound to the inactive proteinase. Translation was initiated at residue Met106. Thus, the recombinant protein represented only the C-terminal domain of the SCP with the additional N-terminal residues 106–113 and C-terminal residues 265 and 266. The structure of the resultant triclinic (type 4) crystals showed almost the complete recombinant protein, from Ala107 (Fig. 2) to the C-terminal residue Ala266, missing only Met106 which presumably was removed in *E. coli*.

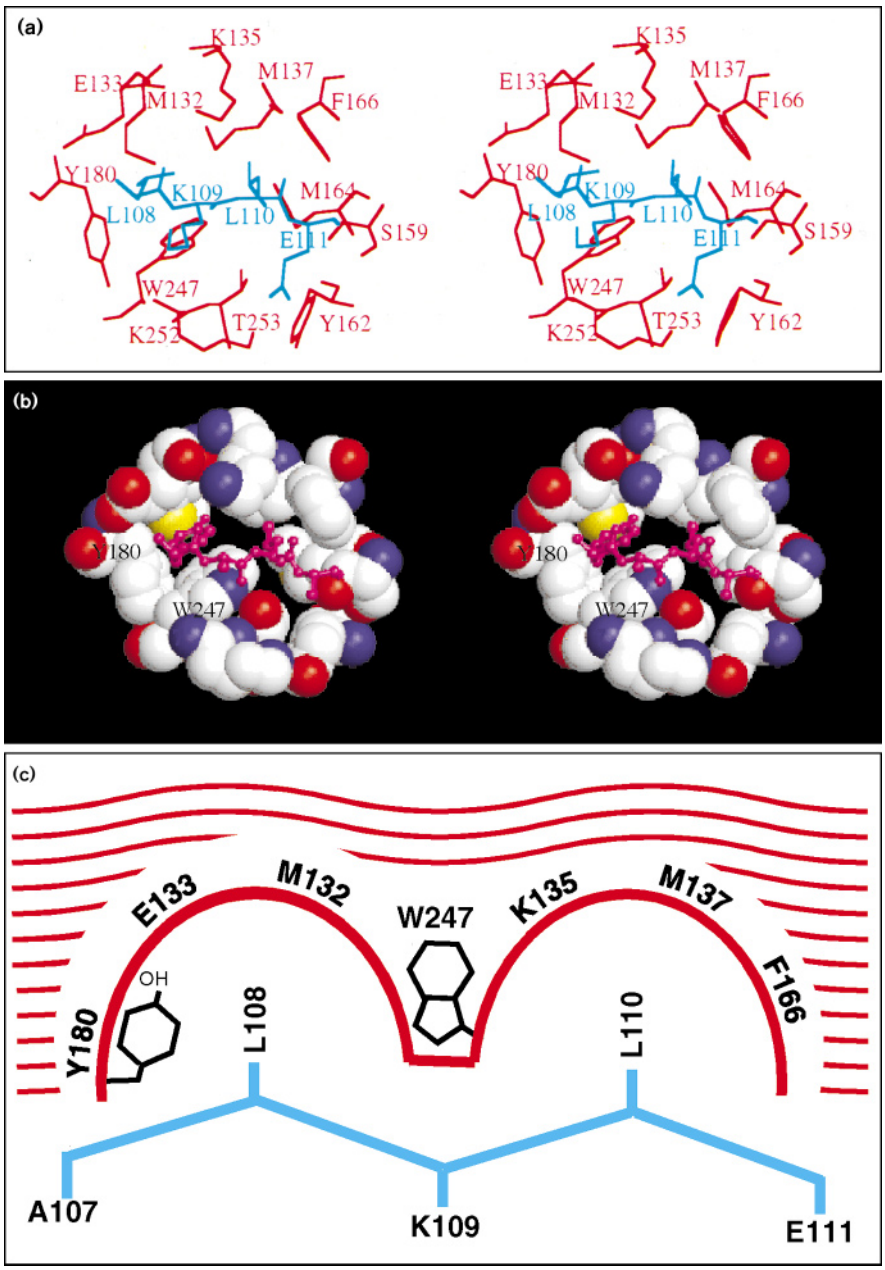
Refinement of the better-resolved 106–266 mutant crystal structure, compared with the previously refined SCP

**Figure 2**




A 2.0 Å resolution ($2F_o–F_c$) electron density map of type 4 SCP crystals showing residues 107–114 in monomer A. Lys109 is only partially visible at the chosen contour level of 1.3σ.

**Figure 3**



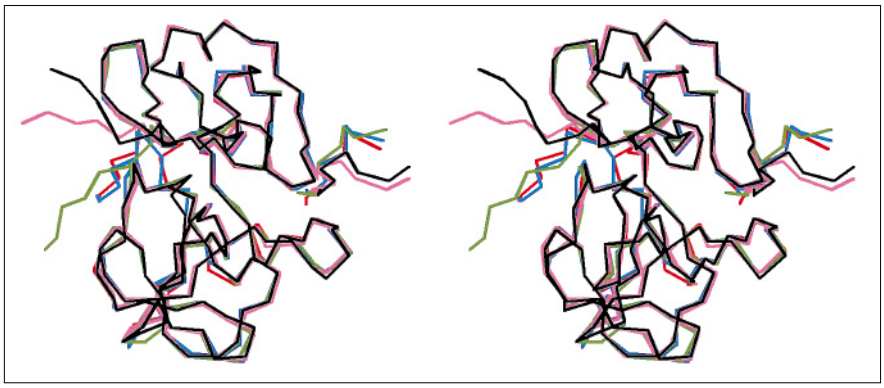

The hydrophobic binding pocket of SCP showing the bound SCP residues 108–111 (the N-terminal arm). **(a)** The pocket in SCP (red) with the bound N-terminal arm residues 108–111 (blue). **(b)** Space filling stereofigure (using standard atom coloring) of the SCP pocket with a model of the E2 residues (400–403), shown in magenta, replacing SCP residues 108–111. The E2 residues are shown as a ball and stick model. **(c)** Side view showing a diagrammatic representation of binding between the hydrophobic pocket and the N-terminal arm residues 108–111. Residues of the capsid hydrophobic pocket are shown in red on the double semicircles. The incoming N-terminal arm residues from an adjacent capsid protein are shown in blue.

structures [23], showed that there had been a one residue frameshift in the placement of 10 consecutive residues in the earlier structure determinations. Subsequent refinement of the corrected structures of both the tetragonal (type 2) and monoclinic (type 3) crystal forms showed that residues 107–113 were not disordered, but their densities had been deleted as a consequence of averaging the molecular structures between type 2 and 3 crystal forms (Table 1) [32]. Residues 114–264 have a similar structure in each form whereas residues 107–113 have different conformations to accommodate the different crystal packing arrangements.

In each of the crystal structures, the N-terminal arms of the SCP were found to bind into a hydrophobic pocket on the surface of neighboring molecules (Fig. 3). The orientation and position of residues 108–111 in the pocket are well conserved, although the approach pathway is vastly different in every case (Fig. 4; Table 1). The pocket is situated between the two β-barrel subdomains of the C-terminal domain, on the same side of the molecule to the proteinase active center (Fig. 5). Residues Leu108 and Leu110 provide the hydrophobic contacts within the pocket. This pocket is made up of two compartments, the first binding Leu108 and the

**Figure 4**

Superposition of the five available structures of SCP. The divergence of their N-termini is seen on the left. Moreover, the N-terminal arm of neighboring molecules converge into the hydrophobic pocket (right). (Superposition was achieved by minimizing the sum of the squared distances between Cα atoms 114–264.)



second binding Leu110, with Trp247 dividing the two sides (Fig. 3c). Leu108 is sandwiched between Tyr180 and Trp247, while Leu110 is between Trp247 and Phe166. The hydrophobic nature of the pocket is well conserved among different alphaviruses (Table 2), with Trp247 being completely conserved and Tyr180 differing only in O'Nyong-nyong virus where it is a phenylalanine. The conserved Glu111 is involved in a salt bridge with Lys252 in SINV, but the latter residue is not conserved among alphaviruses (Table 2). In all alphaviruses, residue 109 is a lysine which points out of the pocket and therefore is irrelevant for binding.

### Biological significance of residues 108–111

Although the N-terminal arm binds into neighboring monomers in various SCP crystal structures, this arm is not of sufficient length to reach from one capsid monomer to another in the tail-to-head interaction seen in the intact virus [1]. Nevertheless, the binding of the capsid protein residues 108–111 into the hydrophobic pocket of the neighboring capsid monomers may have functional significance. Therefore, site-directed mutagenesis of capsid residues 108 and 110, the two residues shown structurally to be important for binding, was carried out to examine their role in virus assembly. These two leucine residues were changed to various combinations of polar residues (Table 3). Viruses rescued from RNA transfections produced plaques that were smaller in size relative to the wild type. Residues Lys109 and Glu111, which are not involved in hydrophobic binding to the pocket, were also mutated. Although these two residues are conserved among alphaviruses (Fig. 1), the virus in which both residues were substituted with alanine exhibited a wild-type plaque phenotype.

Viruses that were mutated at SCP residues Leu108 and Leu110 had reduced growth efficiency, as determined by one-step growth analyses (Table 3). Western blot analysis of cytoplasmic extracts indicated that this deficiency was not a result of translational inhibition or capsid protein degradation (data not shown). Comparative analyses of virus supernatants suggested that the mutants are not defective in their particle to plaque-forming unit (pfu) ratios. Therefore, the virions that assembled were

**Figure 5**

Ribbon diagram of SCP (107–266) showing the active center residues (white) and the hydrophobic pocket residues (red). The bound N-terminal arm peptide is in magenta.

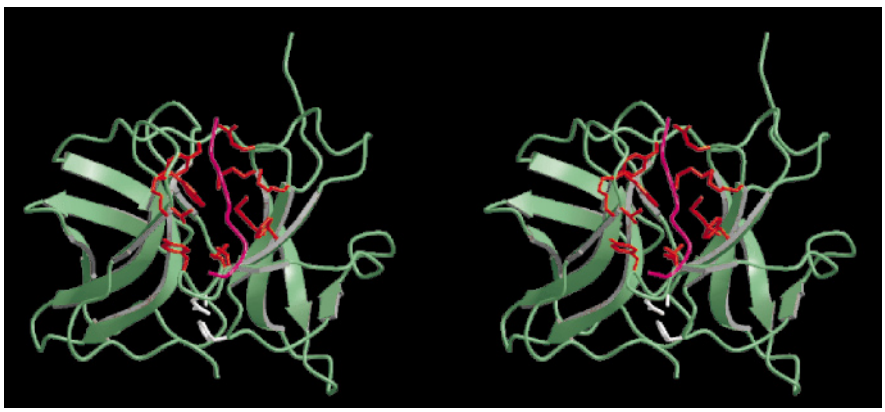


**Table 2**

**Conservation among residues involved in the hydrophobic binding pocket.***

| | SINV (SFV†) residues | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 132 (136) | 133 (137) | 135 (139) | 137 (141) | 159 (163) | 162 (166) | 164 (168) | 166 (170) | 180 (184) | 247 (251) | 252 (255) | 253 (256) |
| SINV | M | E | K | M | S | Y | M | F | Y | W | K | T |
| AURV | M | E | K | I | S | Y | M | F | Y | W | K | T |
| WEEV | V | G | R | M | A | Y | L | Y | Y | W | V | T |
| EEEV | V | G | R | F | A | Y | L | Y | Y | W | V | T |
| VEEV | V | G | K | F | A | Y | L | Y | Y | W | V | T |
| ONNV | V | G | K | M | S | Y | L | C | F | W | I | V |
| RRV | V | G | K | M | S | Y | L | C | Y | W | M | V |
| SFV | V | G | K | M | S | Y | L | C | Y | W | M | V |

*Abbreviations are the same as those of Figure 1. †SFV numbers are given for comparison with the accompanying paper [37].

infectious. Furthermore, a nucleocapsid accumulation assay showed all four mutants failed to produce detectable quantities of nucleocapsid cores in the cytoplasm (Table 3). It is possible that the mutant cores could not withstand the extraction procedure; however, infection of cells with mutant virus Lys109→Ala; Glu111→Ala or wild-type virus resulted in considerable core accumulation.

The first step in assembly is the association of capsid proteins and genomic RNA to form nucleocapsid cores. Since cores accumulate in the cytoplasm of cells infected with wild-type virus, we propose that this first assembly step is fast relative to the incorporation of cores into virus particles. Therefore, the viruses that were mutated at Leu108 and Leu110 have a defect in core assembly such that the rate of nucleocapsid formation is reduced and results in a failure to detect free cores in the cytoplasm. Although the rate of infectious particles released from the cell is also reduced, this decrease is presumably due to the reduction in the rate of nucleocapsid formation and not in that of virus budding.

Residues 97–113 are well conserved among alphaviruses and have previously been implicated in the specific binding of capsid protein to genomic RNA [30,33], as well as in the attachment of the nucleocapsid to ribosomes during disassembly [31]. Residues 108 and 110, discussed above, are within this conserved region and play a role in core formation. Deletion of the upstream residues 97–106 does not affect core production although RNA specificity is disrupted [34]. Unlike wild-type protein, a mutant containing a deletion of capsid residues 97–106 encapsidates the subgenomic mRNA, leading to a higher particle to pfu ratio. Thus, the data taken together are consistent with the mutations at 108 and 110 disrupting binding to the hydrophobic pocket, rather than binding to RNA.

**Table 3**

**Mutations at residues 108–111 of the SCP.**

| Mutation | Plaque phenotype of the virus* | Replication efficiency† | Nucleocapsid accumulation |
|---|---|---|---|
| Toto64 (wild type) | Large | 100% | + |
| Lys109→Ala; Glu111→Ala | Large | 60% | + |
| Leu108→Asn; Leu110→Asp | Medium | 18% | – |
| Leu108→Asn; Leu110→Lys | Medium | 19% | – |
| Leu108→Asp; Leu110→Lys | Medium | 3% | – |
| Leu108→Asp; Leu110→Asn | Medium | 5% | – |

*The plaque phenotype following transfection. †Rate of virus release at 12 h post infections, shown as percent of wild type.

The binding of the N-terminal arm to newly synthesized capsid proteins may be a mechanism by which core formation is initiated. Such associations would tether capsid proteins together prior to RNA binding and assembly. Displacement of the N-terminal arm to facilitate core formation might be accomplished by binding genomic RNA to a region around residues 97–113. Alternatively, if assembly takes place on membranes, the N-terminal arm could be displaced by binding of the E2 cytoplasmic domain to the pocket.

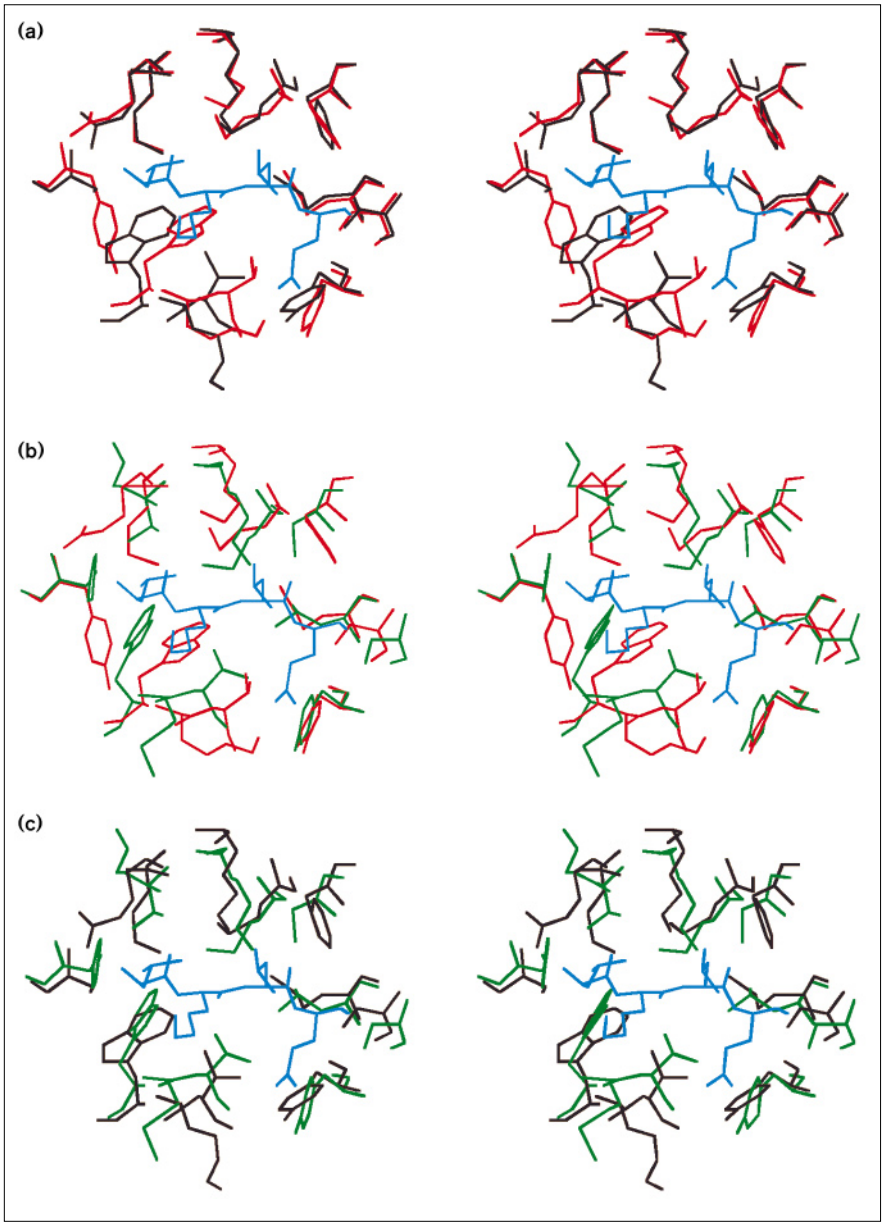
### Binding of E2 to alphavirus capsid proteins

A double mutation in the capsid, Tyr180→Ser; Glu183→Gly [21], allowed virus formation but caused a defect in the uncoating pathway resulting in an increase in the particle to pfu ratio. To investigate the structural basis of the defect, the cDNA containing the double mutation was cloned into the SCP *E. coli* expression vector. Following expression, purification and crystallization of residues 106–264, the structure of the mutant protein was solved (H-KC, *et al.*, and RJK, unpublished data). The mutation Tyr180→Ser alters the structure of the pocket by changing the dihedral angles of the Trp247 side chain such that it fills the space previously occupied by the Tyr180 side chain (Fig. 6a); the shape and hydrophobicity of the pocket are left largely unchanged. Binding of the N-terminal arm is unaffected by the repositioning of the tryptophan. Substitution of Glu183 for glycine has no obvious conformational effect (Glu183 is not within the proximity of the pocket). Therefore, it is likely that the substitution of Tyr180 makes the major contribution to the altered phenotype.

Residues within or near the pocket, in particular Tyr180, have been implicated in the binding of the C-terminal residues of E2 and, hence, in the requirements for correct viral maturation [20,21]. Amino acid residues in E2 have also been identified as being important for budding, and these are located around the completely conserved Tyr400 (Fig. 1) [9]. Mutation of this tyrosine to the hydrophobic residues tryptophan, phenylalanine, leucine or methionine maintains fairly efficient budding, whereas mutation of this residue to the polar residues

**Figure 6**

The hydrophobic pocket of various capsid proteins, with the N-terminal arm shown in blue. The orientation is as in Figure 3a. (a) Wild-type SCP (red) superimposed on the double mutant (Tyr180→Ser; Glu183→Gly) (brown). (b) SFCP (green) superimposed on wild-type SCP (red). (c) SFCP (green) superimposed onto the double mutant (Tyr180→Ser; Glu183→Gly) of SCP (brown).




lysine, serine, asparagine or threonine does not [17]. The hydrophobicity of the E2 Tyr400 residue, the hydrophobic nature of the pocket and the binding of Leu108 and Leu110 into the pocket suggest an analogy between the SCP N-terminal arm residues 108 to 111 and the SINV E2 residues 400 to 403 (Fig. 1). Thus, we propose that the SCP N-terminal arm mimics the binding of E2 to SCP, as modeled in Figure 3b. Such a postulate would also explain why peptides that resemble the appropriate region of E2, when diffused into SCP crystals, were never observable in difference electron density maps (H-KC and MGR, unpublished data); this was presumably because the capsid residues 108–111 were already occupying the pocket. The alignment in Figure 1 suggests that Leu108 is analogous to the conserved Tyr400 of E2. Indeed, substitution of Tyr400 in SFV with leucine produces a budding competent virus [17]. Leu110, the other major residue required for binding, corresponds to a completely conserved Leu402 in the E2 sequences and can be either a leucine or isoleucine in the capsid protein sequences. Residue Lys109, which points away from the pocket and is not involved in binding, is analogous to the E2 residue 401 that is appropriately variable.

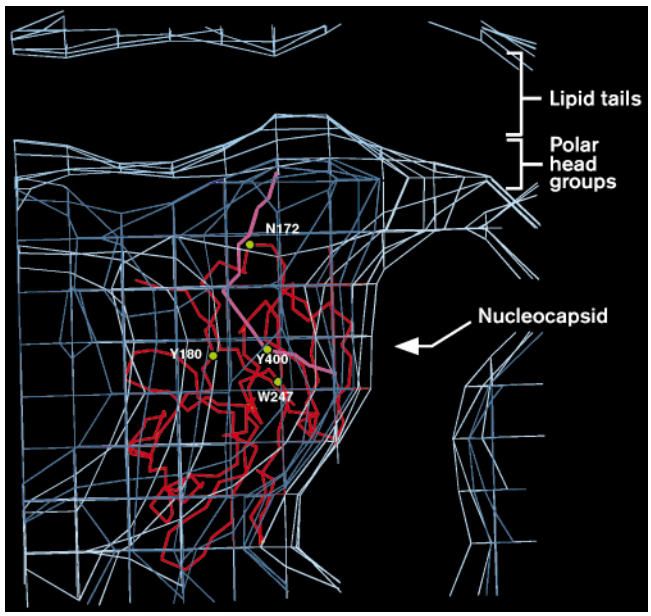
The pocket in the wild-type SFCP structure has greater similarity to the SCP mutant (Tyr180→Ser; Glu183→Gly) structure than it has to wild-type SCP (Fig. 6b,c). Although residue 180 is a tyrosine in both SFCP and SCP, the tyrosine is turned outward in SFCP, creating a space in the pocket similar to that which occurs in the Tyr180→Ser; Glu183→Gly mutant due to the replacement of tyrosine by serine. The resultant space is occupied by Trp247 in both wild-type SFCP and the SCP double mutant. Although the pocket is occupied by residues 108–111 in all SCP structures including the double mutant, the SFCP structures do not show homologous binding. However, the packing organization of the monomers within the SFCP crystals is such that the N-terminal arm is unable to contact any of the binding pockets within the crystals, including that of the same molecule (Table 1). By far, the biggest differences between the SFCP and SCP structures relate to this positioning of Tyr180 and Trp247. Two other differences occur in the vicinity of single amino acid deletions in SFCP relative to SCP. The rms deviation between equivalent Cα atoms is ~0.8 Å.

The location of the hydrophobic pocket as well as the site of the glycoprotein penetration of the viral membrane has been determined by fitting the alphavirus core protein into the cryo-EM electron-density map of the virus [1]. The distance from the site of penetration to the hydrophobic pocket is ~28 Å, which corresponds well with the distance spanned by the ten amino acids of the E2 cytoplasmic domain (residues 391–400). This placement permits binding of E2 into the pocket in a direction consistent with the bound N-terminal arm observed in the crystal structures (Fig. 7). The ten E2 residues in this connecting region are highly conserved among the alphaviruses (Fig. 1) and may therefore also be involved in the positioning of E2 in the hydrophobic pocket. This is further supported by a palmitoylation site at Cys396 [10], confirming the proximity of these residues to the membrane.

### Switch mechanism for alphavirus budding

Coombs and Brown [35] used lactoperoxidase to radioiodinate tyrosines on SINV cores and found that only Tyr180 was available on the core surface. However, Tyr180 as found in the wild-type SCP structures has its phenolic oxygen buried, whereas in SFCP the tyrosine is turned outwards. Furthermore, placement of the SFCP into the RRV cryo-EM density demonstrates that Tyr180 would be exposed on the surface of isolated cores [1]. It is proposed that the structure of SFCP, without the bound N-terminal arm, corresponds to that in the cytoplasmic alphavirus cores, whereas the structures of SCP with the bound N-terminal arm represent the core structure with E2 bound, as in virions [36] (Fig. 8a).

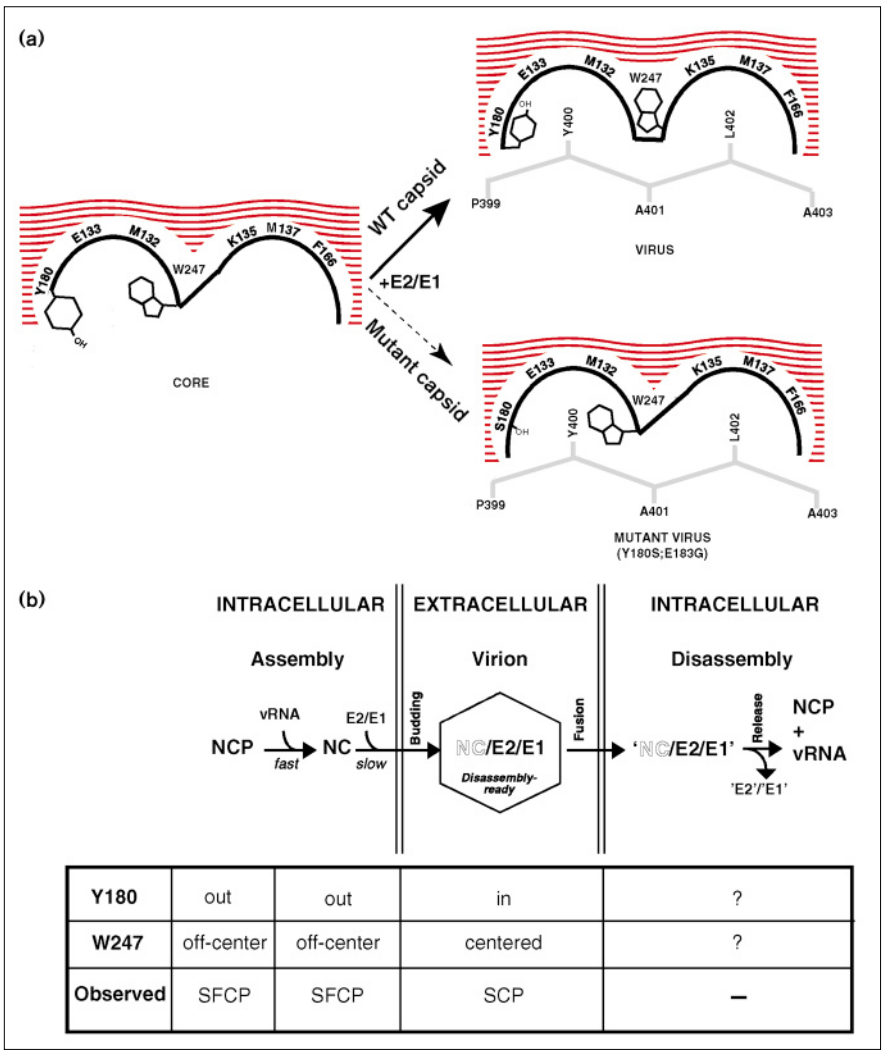
**Figure 7**




Cross section of the cryo-EM density for RRV showing the SCP Cα structure (red) fitted to the nucleocapsid. The lipid bilayer membrane is at the top (light blue density) with its pinched region corresponding to the penetration site of the glycoprotein spike. The E2 cytoplasmic domain (magenta) is shown as an extended polypeptide from residue 391 near the membrane to the hydrophobic pocket (of which Tyr180 and Trp247 are indicated) which binds residues 400–403. No attempt was made to fit the polypeptide to the low resolution density. This model building demonstrates only that the path length is reasonable and that the polypeptide can be placed into the pocket in the correct orientation. The rest of the sequence (404–423) is also hydrophobic with Cys416 and Cys417 being palmitoylated and again associated with the membrane (not modeled). Two contours of electron density are shown (light and dark blue).

The temperature-sensitive Tyr180→Ser; Glu183→Gly mutant is defective in the entry-uncoating process although virus fusion is unaffected. The maximum production of infectious virus occurs between 30 and 40°C [21]. Thus, the crystal structure at room temperature may represent the protein in its non-permissive state (Fig. 6). Although the double mutant SCP binds the N-terminal arm in the crystal structure, the absence of a tyrosine at position 180 prevents the movement of Trp247 to the position seen in the wild-type SCP structure. Binding of E2 to the mutant hydrophobic pocket would also fail to change the conformation of the hydrophobic pocket. Therefore the double mutant probably lacks the ability to form the optimal capsid–E2 interactions at the non-permissive temperature, which manifests itself in the disassembly of the virus (Fig. 8a). Although the ectodomains of E1 and E2 of the mutant virus are competent for cell attachment and viral fusion, the interaction of the nucleocapsid core and the E2 cytoplasmic domain is adversely affected (Fig. 8b).

**Figure 8**




| | | | | |
|---|---|---|---|---|
| **Y180** | out | out | in | ? |
| **W247** | off-center | off-center | centered | ? |
| **Observed** | SFCP | SFCP | SCP | – |

The switch mechanism and its role in virus assembly and disassembly. (a) Diagrammatic representation of the switch that occurs within the capsid hydrophobic pocket upon binding of the E2 C-terminal domain. Residues of the capsid hydrophobic pocket are shown on the double semicircles. The incoming E2 C-terminal domain residues are shown in grey. The pocket on the left represents the unbound (no E2) state of capsid protein, modeled on the basis of the SFCP structure. On the basis of SCP structures it is hypothesized that upon binding of E2 to the wild-type capsid the pocket undergoes a conformational change (upper right). The capsid mutant Tyr180→Ser; Glu183→Gly, is still capable of binding E2 and the virus can bud from cells, but the conformational change fails to occur due to the absence of Tyr180 (lower right). (b) Diagrammatic representation of the virus assembly and disassembly pathways. Assembly of the capsid protein (NCP) into cores (NC) occurs rapidly in the presence of viral RNA and with no observed intermediates. The association of cores with envelope glycoproteins E1 and E2 occurs in the next step at a slower rate, with a concomitant conformational change in the hydrophobic pocket which activates the core for disassembly. The virions that have budded from the cell have undergone a conformational change in the core ('disassembly-ready') required for optimal disassembly at a later time. Following fusion of the viral and cellular membranes, the nucleocapsid core is released into the cytoplasm and rapidly falls apart. The Tyr180→Ser; Glu183→Gly double mutant is either deficient in the release of the core into the cytoplasm or in uncoating. The table describes the cystallographically observed states of the capsid protein and hydrophobic pocket in relation to the proposed assembly/disassembly pathway.

## Biological implications

The alphaviruses are a group of enveloped positive-strand RNA viruses that are transmitted by arthropod vectors. They are responsible for a variety of human illnesses including encephalitis and polyarthritis and are represented by such members as Sindbis virus, Semliki Forest virus and Ross River virus. Structural studies have determined that the 690 Å diameter virus particle has an outer protein shell composed of two transmembrane glycoproteins, E1 and E2, arranged into surface spikes. Three pairs of the E1/E2 heterodimer constitute a spike and penetrate a host-derived lipid bilayer that envelopes a nucleocapsid core. This inner core is assembled by the interaction of the genome RNA with 240 copies of capsid protein arranged as a $T$=4 icosahedron. A short stretch of amino acids found on the cytoplasmic face of the E2 transmembrane glycoprotein mediates the interaction between the outer glycoprotein spikes and the inner capsid core. This interaction is responsible for the control of virus budding from the membrane of infected cells.

We have shown that a hydrophobic pocket on the surface of the Sindbis virus core protein (SCP) binds the N-terminal arm (residues 108–111) of neighboring molecules in crystal lattices. Mutational analysis of SCP residues 108 and 110 suggests a role for these residues in the assembly of nucleocapsid cores, possibly by associating capsid proteins in a manner similar to that observed in crystal lattices. Presumably, such an association would facilitate the assembly process of the cores.

Analogy between the amino acid sequence of the SCP N-terminal arm and the cytoplasmic domain of the E2 glycoprotein suggests an interaction of Tyr400 and Leu402 of E2 with the hydrophobic pocket on the outer surface of the nucleocapsid protein. The mutational analysis of Semliki Forest virus core protein (SFCP)

residues Tyr180, Trp247 and Val132 (valine is replaced by methionine in SCP), which are found in the hydrophobic pocket [37], is consistent with these residues playing an important role in interacting with the E2 cytoplasmic domain. It also provides a genetic basis to support the structural studies presented.

It is proposed that the mechanism by which the cores associate with the E2 glycoproteins is regulated by a switch between two conformations of the hydrophobic pocket. This switch causes a conformational change in the nucleocapsid such that the core is readied for disassembly. Without the optimal core–E2 interaction (as in the double mutant Tyr180→Ser; Glu183→Gly), the core does not undergo the conformational changes required for subsequent disassembly. Insertion of the E2 cytoplasmic residues into the hydrophobic pocket, with the concomitant burial of Tyr180, should stabilize the assembled virion. This mechanism is similar to that found in the rhino and enteroviruses, in which there is a cellular hydrophobic 'pocket factor' which is thought to stabilize virions in the extracellular environment prior to receptor binding. The pocket factor is released on viral attachment to the cell, thus destabilizing the virus in readiness for uncoating [38]. Burial of the aromatic residues into the hydrophobic pocket is analogous to insertion of the pocket factor into some picornaviruses.

The driving force for budding is provided by making the hydrophobic surfaces of the capsid protein and the E2 C-terminal residues inaccessible to water. Burial of the surface area of the hydrophobic pocket covered by E2 (316 Å$^2$) corresponds to an energy of 7.6 kcal M$^{-1}$ [39]. This area is buried 240 times in the core surface during the budding process and this would provide significant energy for this last step in virus assembly. Similar mechanisms may also drive the budding of other enveloped viruses.

## Materials and methods

### *Crystallization and structure determination of alphavirus capsid proteins*

Full-length SCP and SFCP were isolated from virions as previously described [22,24]. Truncated forms of SCP were expressed in an *E. coli* expression system (H-KC, SL, Y-P Zhang, BR McKinney, GW, MGR and RJK, unpublished data). The recombinant mutant protein SCP 106–266 (Ser215→Ala) was crystallized into space group P1 (a=29.0 Å, b=56.5 Å, c=60.8 Å, $\alpha$=93.1°, $\beta$=96.7°, $\gamma$=94.9°) with two independent molecules per unit cell. The diffraction of these type 4 crystals (Table 4) extended to a much higher resolution and was of better quality than all other previously studied SCP or SFCP crystals. The structure was solved using the molecular replacement method and then refined to an R-factor of 18.4% ($R_{free}$=27.99) with an rms deviation of bond lengths from their idealized values of 0.011 Å. There were 133 water molecules included in the structure. There were two SCP molecules in the triclinic cell related to each other by 14.2° rotation and packed in a head-to-tail fashion similar to, but not the same as, that which occurs in the nucleocapsid cores (Table 1).

**Table 4**

**X-ray diffraction data for type 4 crystals.**

| Resolution (Å) | $R_{sym}$* | % of observed data |
|---|---|---|
| ∞–3.91 | 2.4 | 95.6 |
| 3.91–3.11 | 3.4 | 95.7 |
| 3.11–2.71 | 4.8 | 92.2 |
| 2.71–2.47 | 5.8 | 88.7 |
| 2.47–2.29 | 6.9 | 76.1 |
| 2.29–2.15 | 9.8 | 41.7 |
| 2.15–2.00 | 9.8 | 20.3 |

Reflections with $I_{hi} < \sigma(I_h)$ were rejected.
*$R_{sym}=[\sum_h\sum_i|(I_{hi}-<I_h>)|/\sum_h\sum_i I_{hi}]\times 100$.

Refinement of the better-resolved type 4 SCP structure, compared to the earlier determined SCP type 2 and 3 crystal structures, showed a frameshift in the positioning of the amino acids between residues 250 and 260. In addition, the C-terminal Trp264 side-chain placement was rotated by about 90°. The more recently determined structures of SFCP type I and II crystal forms (H-KC, GL, SL, GW and MGR, unpublished data) were like the type 4 SCP structure, but not like the earlier SCP structures with respect to residues 250–260 and residue 264.

### *Re-refinement of the wild-type SCP structures*

Because of the differences in the type 4 SCP compared with the older, less resolved SCP structures, the previous results were re-examined. Diffraction data were re-collected for types 2 and 3 SCP crystal forms (H-KC, SL, Y-P Zhang, BR McKinney, GW, MGR and RJK, unpublished data) to a higher resolution than previously available [23]. Omit maps showed that the earlier SCP structures had been in error between residues 250–260 as well as in the orientation of the C-terminal Trp264. In addition, aided by the correct interpretation and better data, the N-terminal arms could now be traced between residues 107–113.

### *Structures of the truncated SCP (106–264) and truncated double mutant (Tyr180→Ser; Glu183→Gly)*

Wild-type SCP from 106–264 was expressed in *E. coli* (H-KC, SL, Y-P Zhang, BR McKinney, GW, MGR and RJK, unpublished data) and crystallized isomorphously with the wild-type SCP type 2 crystals. The cDNA from the double mutant used by Lee and Brown [21], in which Tyr180 had been changed to serine and Glu183 to glycine (Tyr180→Ser; Glu183→Gly), was introduced into the *E. coli* expression system. Following expression and purification, the protein crystallized isomorphously with type 2 wild-type crystals.

### *Isolation and characterization of mutant viruses*

Double substitutions in the N-terminal arm of the capsid protein were constructed using oligonucleotide-directed mutagenesis as previously described [40]. The parental virus, Toto64, is described elsewhere and produces wild-type virus [34]. Following identification of the full-length mutant cDNA clones, infectious RNAs were transcribed *in vitro* and transfected into baby hamster kidney (BHK) cells. Mutant viruses were rescued by plaque purification and stock viruses were generated on BHK cell monolayers.

One step growth curve analyses were carried out in BHK cells at 37°C as previously described except that a multiplicity of infection of 1 was used to infect cells [40]. Results are given for the replication efficiency of the mutant viruses as compared with the wild-type virus at 12 h post infection. Following virus infection of BHK cells, nucleocapsid accumulation assays were carried out as previously described [18].

### *Accession numbers*

Coordinates have been deposited with the Brookhaven Protein Data Bank, accession number 1SVP.

## Acknowledgements

We wish to thank Bonnie McKinney and Zhu Zhu for excellent technical assistance and Peter Liljeström for providing us with his data prior to publication. This work was supported by an NIH program project grant (AI35212) to the Structural Virology group at Purdue University, including MGR and RJK, by an NSF grant (MCB-9102855) to MGR and an NIH grant (AI33982) to RJK. DTB also acknowledges the support of an NIH grant (AI14710). KEO is a recipient of a Purdue Research Foundation fellowship.

## References

1. Cheng, R.H., *et al*., & Baker, T.S. (1995). Three-dimensional structure of an enveloped alphavirus with *T*=4 icosahedral symmetry. *Cell* 80, 1–20.
2. Fuller, S.D., Berriman, J.A., Butcher, S.J. & Gowen, B.E. (1995). Low pH induces swiveling of the glycoprotein heterodimers in the Semliki Forest virus spike complex. *Cell* 81, 715–725.
3. Paredes, A.M., *et al*., & Prasad, B.V.V. (1993). Three-dimensional structure of a membrane-containing virus. *Proc. Natl. Acad. Sci. USA* 90, 9095–9099.
4. Smith, T.J., *et al*., & Baker, T.S. (1995). Putative receptor binding sites on alphaviruses as visualized by cryoelectron microscopy. *Proc. Natl. Acad. Sci. USA* 92, 10648–10652.
5. Strauss, J.H. & Strauss, E.G. (1994). The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* 58, 491–562.
6. Zhao, H. & Garoff, H. (1992). Role of cell surface spikes in alphavirus budding. *J. Virol.* 66, 7089–7095.
7. Strauss, J.H., Strauss, E.G. & Kuhn, R.J. (1995). Budding of alphaviruses. *Trends Microbiol.* 3, 346–350.
8. Liljeström, P. & Garoff, H. (1991). Internally located cleavable signal sequences direct the formation of Semliki Forest virus membrane proteins from a polyprotein precursor. *J. Virol.* 65, 147–154.
9. Gaedigk-Nitschko, K. & Schlesinger, M.J. (1991). Site-directed mutations in Sindbis virus E2 glycoprotein's cytoplasmic domain and the 6K protein lead to similar defects in virus assembly and budding. *Virology* 183, 206–214.
10. Ivanova, L. & Schlesinger, M.J. (1993). Site-directed mutations in the Sindbis virus E2 glycoprotein identify palmitoylation sites and affect virus budding. *J. Virol.* 67, 2546–2551.
11. Liu, N. & Brown, D. T. (1993). Phosphorylation and dephosphorylation events play critical roles in Sindbis virus maturation. *Virology* 196, 703–711.
12. Liu, N. & Brown, D.T. (1993). Transient translocation of the cytoplasmic (Endo) domain of a type I membrane glycoprotein into cellular membranes. *J. Cell Biol.* 120, 877–883.
13. Barth, B.U., Suomalainen, M., Liljeström, P. & Garoff, H. (1992). Alphavirus assembly and entry: role of the cytoplasmic tail of the E1 spike subunit. *J. Virol.* 66, 7560–7564.
14. Metsikkö, K. & Garoff, H. (1990). 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.
15. Suomalainen, M., Liljeström, P. & Garoff, H. (1992). Spike protein–nucleocapsid interactions drive the budding of alphaviruses. *J. Virol.* 66, 4737–4747.
16. Kail, M., *et al*., & Vaux, D. (1991). The cytoplasmic domain of alphavirus E2 glycoprotein contains a short linear recognition signal required for viral budding. *EMBO J.* 10, 2343–2351.
17. Zhao, H., Lindqvist, B., Garoff, H., von Bonsdorf, C.H. & Liljeström, P. (1994). A tyrosine-based motif in the cytoplasmic domain of the alphavirus envelope protein is essential for budding. *EMBO J.* 13, 4204–4211.
18. Lopez, S., Yao, J.S., Kuhn, R.J., Strauss, E.G. & Strauss, J.H. (1994). Nucleocapsid–glycoprotein interactions required for assembly of alphaviruses. *J. Virol.* 68, 1316–1323.
19. Collier, N.C., Adams, S.P., Weingarten, H. & Schlesinger, M.J. (1992). Inhibition of enveloped RNA virus formation by peptides corresponding to glycoprotein sequences. *Antivir. Chem. Chemother.* 3, 31–36.
20. Lee, H., Ricker, P.D. & Brown, D.T. (1994). The configuration of Sindbis virus envelope proteins is stabilized by the nucleocapsid protein. *Virology* 204, 471–474.
21. Lee, H. & Brown, D.T. (1994). Mutations in an exposed domain of Sindbis virus capsid protein result in the production of noninfectious virions and morphological variants. *Virology* 202, 390–400.
22. Choi, H.K., *et al*., & Wengler, G. (1991). Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature* 354, 37–43.
23. Tong, L., Wengler, G. & Rossmann, M. G. (1993). Refined structure of Sindbis virus core protein and comparison with other chymotrypsin-like serine proteinase structures. *J. Mol. Biol.* 230, 228–247.
24. Boege, U., Wengler, G., Wengler, G. & Wittmann-Liebold, B. (1981). Primary structures of the core proteins of the alphaviruses Semliki Forest virus and Sindbis virus. *Virology* 113, 293–303.
25. Hahn, C.S., Strauss, E.G. & Strauss, J.H. (1985). Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid protein autoprotease. *Proc. Natl. Acad. Sci. USA* 82, 4648–4652.
26. Melancon, P. & Garoff, H. (1987). Processing of the Semliki Forest virus structural polyprotein: role of the capsid protease. *J. Virol.* 61, 1301–1309.
27. Hahn, C.S. & Strauss, J.H. (1990). Site-directed mutagenesis of the proposed catalytic amino acids of the Sindbis virus capsid protein autoprotease. *J. Virol.* 64, 3069–3073.
28. Rossmann, M.G. & Johnson, J.E. (1989). Icosahedral RNA virus structure. *Annu. Rev. Biochem.* 58, 533–573.
29. Weiss, B., Geigenmüller-Gnirke, U. & Schlesinger, S. (1994). Interactions between Sindbis virus RNAs and a 68 amino acid derivative of the viral capsid protein further defines the capsid binding site. *Nucleic Acids Res.* 22, 780–786.
30. Weiss, B., Nitschko, H., Ghattas, I., Wright, R. & Schlesinger, S. (1989). Evidence for specificity in the encapsidation of Sindbis virus RNAs. *J. Virol.* 63, 5310–5318.
31. Wengler, G., Würkner, D. & Wengler, G. (1992). Identification of a sequence element in the alphavirus core protein which mediates interaction of cores with ribosomes and the disassembly of cores. *Virology* 191, 880–888.
32. Tong, L., Choi, H.K., Minor, W. & Rossmann, M.G. (1992). The structure determination of Sindbis virus core protein using isomorphous replacement and molecular replacement averaging between two crystal forms. *Acta Cryst. A* 48, 430–442.
33. Geigenmüller-Gnirke, U., Nitschko, H. & Schlesinger, S. (1993). Deletion analysis of the capsid protein of Sindbis virus: identificaton of the RNA binding region. *J. Virol.* 67, 1620–1626.
34. Owen, K. E. & Kuhn, R. J. (1996). Identification of a region in the Sindbis virus nucleocapsid protein that is involved in specificity of RNA encapsidation. *J. Virol.*, 70, 2757–2763.
35. Coombs, K. & Brown, D. T. (1987). Topological organization of Sindbis virus capsid protein in isolated nucleocapsids. *Virus Res.* 7, 131–149.
36. Coombs, K., Brown, B. & Brown, D. T. (1984). Evidence for a change in capsid morphology during Sindbis virus envelopment. *Virus Res.* 1, 297–302.
37. Skoging, U., Vihinen, M., Nilsson, L. & Liljeström, P. (1996). Aromatic interactions define the binding of the alphavirus spike to its nucleocapsid. *Structure*, 4, 519–529.
38. Rossmann, M.G. (1994). Viral cell recognition and entry. *Protein Sci.* 3, 1712–1725.
39. Chothia, C. (1975). Structural invariants in protein folding. *Nature* 254, 304–308.
40. Kuhn, R. J., Hong, Z. & Strauss, J. H. (1990). Mutagenesis of the 3′ nontranslated region of Sindbis virus RNA. *J. Virol.* 64, 1465–1476.
41. Rümenapf, T., Strauss, E.G., & Strauss, J.H. (1995). Aura virus is a new world representative of Sindbis-like viruses. *Virology* 208, 621–633.