The structure of simian virus 40 refined at 3.1 Å resolution
Thilo Stehle, Steven J Gamblin†, Youwei Yan‡ and Stephen C Harrison*

Background: The structure of simian virus 40 (SV40), previously determined at 3.8 Å resolution, shows how its pentameric VP1 assembly units are tied together by extended C-terminal arms. In order to define more precisely the possible assembly mechanisms, we have refined the structure at 3.1 Å resolution.

Results: New data from a high-intensity synchrotron source have been used for phase extension by electron-density averaging and refinement, exploiting only the strict 5-fold non-crystallographic symmetry for the real-space averaging steps. The accurate model enables us to study important structural features of the virus particle in detail. The remarkably invariant core of the VP1 pentamer bears the docking sites for the C-terminal arms from other pentamers. These contacts are the principal way in which pentameric assembly units are linked together in the capsid. Only at the interface between five-coordinated and six-coordinated pentamers do the pentamer cores appear to interact strongly. There are two cation-binding sites per VP1 monomer, seen in a soaking experiment with gadolinium nitrate. These sites are quite close to each other at the interfaces between pentamers.

Conclusions: We propose that the contact between five-coordinated and six-coordinated pentamers may help to generate a six-pentamer nucleus, with which further pentamers can assemble to generate the complete particle. Calcium ions probably stabilize the structure of the assembled particle, rather than direct its assembly.

Introduction
Simian virus 40 (SV40) and murine polyomavirus (polyoma) are the largest virus particles whose structures have so far been determined at the atomic level [1–3]. They are composed of 72 pentamers of the major structural protein, VP1, centered on the vertices of a T=7d icosahedral lattice; a single copy of an internal protein (VP2 or VP3) is associated with each pentamer. The core of each VP1 subunit is a β-roll domain. Five of these domains pack tightly to form each pentamer, with a very extensive surface area buried between them. The cores of the 12 pentamers centered on strict 5-fold axes are of identical structure to those of the 60 pentamers with only local 5-fold symmetry. The last 45–50 residues at the C terminus of each VP1 form an extended arm, which interacts extensively with subunits of another pentamer, acting to tie the virus particle together. Details of the initial structure determination and a low-resolution comparison of the two virus particles are presented in the preceding paper, along with a general introduction to their composition and overall molecular architecture [3]. Refined data collection techniques and technical advances in several areas have enabled us to collect a new data set from SV40 crystals diffracting to spacings beyond 3.1 Å. Using these data and the previous 3.8 Å model, we have carried out phase extension through 5-fold 'icosahedral' real-space electron-density averaging [4] and crystallographic refinement [5].

The resulting structure, presented in this report, enables us to describe the virus capsid in almost atomic-level detail and allows us to pose questions about events in the life cycle of the virus.

Results
Notation
A schematic representation of the virus architecture is shown in Figure 1. The shell is constructed of 72 VP1 pentamers, which are tied together by their C-terminal arms. Twelve of the pentamers are surrounded by five other pentamers, and sixty are surrounded by six pentamers [1,6]. Each pentamer donates five arms to and receives five arms from surrounding pentamers. The following nomenclature, which is essentially the same as that of Liddington et al. [1], is used throughout the text. A 'monomer' refers to one VP1 molecule. The 12 pentacoordinated pentamers, centered on the icosahedral 5-fold axes, conform to strict icosahedral symmetry and are called 'strict pentamers'. The 60 hexacoordinated pentamers, at positions of local 5-fold symmetry, are referred to as 'local pentamers'. The VP1 monomers in six different bonding environments are named α, α′, α″, β, β′ and γ. The secondary structural elements of a monomer are labeled as follows: β-strands — A, B, C, etc.; helices — 3A, 3B, 3C (310-helices) and αC (α-helix); loops — AB, BC1, BC2, CD, etc. In order to retain the general numbering scheme...
of the β-strands introduced by Liddington et al. [1], which is consistent with the nomenclature of other viruses sharing the ‘jelly-roll’ motif, the strands in the EF-loop are labeled E', E'' and E'''". All elements within a reference monomer are shown without a suffix. Elements which belong to a different pentamer are underlined (e.g. strand J or residue Phe355). Elements in different monomers of the same pentamer carry a superscripted (+) if they belong to the clockwise neighbor (viewed from the outside of the virus) in the reference pentamer (e.g. residue Cys104+, strand F+) or a superscripted (-) if they belong to the anticlockwise neighbor (e.g. residue Cys104-, strand F-).

Accuracy of the model
We have used several means of assessing our model and of estimating the errors in the coordinates. Firstly, the model is in good agreement with the observed diffraction data (R-factor of 25.4% for all data in the range 12–3.1 Å) as well as with stereochemical requirements (root mean square deviations [rmsds] from ideality of bond lengths and bond angles are 0.02 Å and 2.1°, respectively). Secondly, the final map, calculated with 2Fo-Fc coefficients and phases derived from the last round of averaging at 3.1 Å resolution, is very detailed and shows the orientation and conformation of most side chains and peptide bonds. The model is in good agreement with this map, as shown by the real-space density correlation coefficients calculated in O [7] (Fig. 2a). The final map is also very consistent with the essentially unbiased map obtained through phase extension from spacings of 5 Å to 3.35 Å, using only the calculated phases for the extension steps. Thirdly, because only 5-fold icosahedral non-crystallographic symmetry (NCS) was used during averaging and refinement, a comparison of the six independently refined VP1 monomers in one icosahedral asymmetric unit allows us to assess the coordinate error. The structurally similar parts of all six VP1 monomers (residues 21–97 and 111–297) can be superimposed on each other with rmsds of 0.3–0.5 Å (Ca atoms only) and 0.5–0.7 Å (all atoms). The distances between the Ca atoms of the superimposed monomers (Fig. 2b) are in most cases less than 0.5 Å. The temperature factor (B-factor) plot (Fig. 2c) also shows that the thermal parameters for similar residues in different VP1 monomers agree very well with each other. Finally, a Ramachandran plot [8] confirms that the values for the dihedral angles in the protein backbone fall well within the expected regions (data not shown). About 75% of the residues are found in the ‘most favored’ regions, as calculated with PROCHECK [9], and none are in regions forbidden for protein residues. Residues at the edges of the allowed regions are generally found either in special ‘strained’ environments, for example in β-turns or next to proline residues, or in very mobile regions of the protein.
We note three regions in which the electron density is poorly defined. The first encompasses the N-terminal 13 amino acid residues, which are not visible at all. This part of the monomer is probably disordered, and it is not contained in our model. It has a number of positively charged residues and prolines. The second region is the poorly defined CD-loop. Lastly, in only three of the six independent VP1 monomers (α', β' and γ) do the C termini form ordered hairpin loops. The C-terminal 6 residues of VP1 monomers α and α' and the C-terminal 18 residues of β are disordered and excluded from the model presented here. The larger disordered portion of monomer β is explained by the location of these residues at the pentamer–pentamer interface between monomers γ (Fig. 1b). At this interface, the pentamers approach each other too closely to leave space for ordered β C termini. For the same reason, the C-terminal arm of monomer γ is poorly defined between residues 301γ and 320γ, which also reside at the γ–γ contact.

**Temperature factors**

A plot of temperature factors versus residue number (Fig. 2c) suggests very low mobility in the core of VP1, especially in regions of secondary structure, but significantly higher mobility in loop regions. The variations in temperature factor across the molecule are quite dramatic. Regions with extremely high thermal factors (B>120 Å²) are found exclusively at the base of VP1, on the inner side of the virion shell. This region includes the N termini, the CD-loops and the C-terminal arms. Thus, the temperature factor values fall into three very distinct zones within a pentamer: the outer loop region with relatively high values, the core (β-sheet) region with low values, and the inward-facing surface of the pentamer, with extremely high temperature factors. At this resolution, the structure does not allow a distinction between static and dynamic disorder. It is likely that interactions with the internal viral proteins (VP2 and VP3) and/or the DNA, which do not obey the icosahedral symmetry, contribute to the apparent disorder on the inside surface of the capsid by forming non-symmetrical contacts.

**The jelly-roll β-barrel**

A Ca trace of a VP1 monomer of SV40 is shown in Figure 3a, and a schematic drawing is shown in Figure 3b. The dominant secondary structural elements are β-strands, which are folded into a jelly-roll β-barrel. These are shown schematically in Figure 4. The barrel has a six-stranded sheet (strands ΑΒIDG2) and a five-stranded sheet (strands CHEFG1'). The major parts of these two sheets correspond to the 'BIDG' and 'CHEF' sheets seen in other virus structures [10,11], but each is augmented here by contributions from other monomers. The five-stranded CHEFG1' sheet has a regular hydrogen-bond pattern and twist. Four strands belong to one monomer, and strand G1'-at the inner edge of the sheet comes from a neighboring VP1 molecule in the same pentamer. In the six-stranded ΑΒIDG2 sheet, the invading C-terminal arm of another
The structure of VP1. (a) Stereo Ca trace of a VP1 monomer (residues 14–316, thin lines) and the C-terminal arm of another pentamer that invades it (residues 268–361, thick lines). (Figure generated with MOLSCRIPT [66].) (b) Ribbon drawing of a VP1 subunit of SV40. Strands E' and G1', shown in gray, belong to neighboring monomers within one pentamer. The invading arm (shown in magenta) belongs to a neighboring pentamer. The refined structure shows an additional short \(3_{10}\)-helix, 3C, which was not seen in the unrefined structure [1]. In order to retain the original nomenclature, the C-terminal helix is named \(\alpha\)C. (Figure generated with RIBBONS [67].)

The loops that connect the various \(\beta\)-strands of the core structure are quite extensive at the outward-facing surface and relatively short at the inward-facing end of the pentamer. An exception is the EF-loop, which emanates from the ‘bottom’ of the subunit and contains 59 residues. Some of these residues form a three-stranded antiparallel sheet (strands E', E'' and E'''). This sheet lies against the CHEFG1' surface of the jelly-roll. Strands E' and E'' are connected by a short \(3_{10}\)-helix (3C). The CD-loop, also at the base of the pentamer, is described in greater detail below. At the outward-facing end, the BC-loop can be subdivided into two distinct smaller loops, denoted BC1 (residues 52–66) and BC2 (residues 67–88). The DE-loop contains a short \(3_{10}\)-helix (3B). The HI-loop is actually an antiparallel \(\beta\)-ribbon; its strands are labelled H' and I'.
Comparison with jelly-roll domains in other viruses

Many other viral coat proteins have jelly-roll domains [10–25]. A close inspection of the jelly rolls of SV40, tomato bushy stunt virus (TBSV) [12] and the adenovirus hexon [25] shows that strand D of the BIDG sheet contains a β-bulge at a similar position in all cases. In SV40, this bulge is at residues 113/114 (see Fig. 4). The bulge can be used as an initial guide for the superposition of the SV40 barrel onto those of TBSV and the adenovirus hexon based on the BIDG sheet. The result of superimposing the jelly-roll domains in this manner is shown in Figure 5. In all cases, the BIDG sheets align reasonably well. The CHEF sheets of TBSV and SV40 are, however, in very different positions and orientations (Fig. 5a). That is, the packing of the two sheets, BIDG and CHEF, against each other is quite distinct in the two viruses. The β-barrel of SV40 is in somewhat better agreement with the two β-barrels of the adenovirus hexon, and indeed the SV40 barrel aligns as well with either of the hexon β-barrel domains as the latter do with each other (Fig. 5b).

The pentamer core

The ‘core’ of a VP1 pentamer is formed by the tight packing of the jelly-roll β-barrels around a 5-fold axis. Each monomer forms extensive contacts with its neighbors, as shown in Figure 6. The surface buried by the contact is 2800 Å². The shape of the jelly-roll domain and the interactions of its loops are such that the pentamer is essentially cylindrical in outer aspect, with a conical hole along its axis opening out towards the interior of the virus. The most striking interaction is through β-strand G, which is shared between two monomers. At the end of strand F, the polypeptide chain loops across to form hydrogen bonds (as G1) along the edge of the (CHEF)² sheet of the clockwise neighbor. The chain then follows a sharp, rectangular bend and forms the G2 strand in its ‘parent’ monomer (see Fig. 3b). The hairpin loop within strand G2 protrudes markedly and approaches the G1 strand of the anticlockwise neighbor, which has to undergo the sharp bend seen between G1 and G2 to avoid a collision. The hairpin loop is conserved in polyoma. It is conceivable that this loop plays a role in pentamer formation by guiding the G-strand of each monomer towards the BIDG2 sheet of a neighboring VP1 monomer.

An even more extensive contact, as judged by buried surface area, is mediated by the DE-loop, which protrudes markedly from one monomer and inserts deeply into a cleft between the CHEF sheet and the EF-loop of its clockwise neighbor (Fig. 6). This interaction completely shields the upper part of the AJBIDG2 sheet from solvent. The DE-loop and especially helix 3B have very few contacts with their own monomer.
On the outer surface of the virus, the tip of loop DE packs against loop HI*, which in turn packs against the tip of loop DE*. The tip of loop BC2 contacts loops DE* and BC1*. In this way, the tightly packed interface extends to the virus surface, and the interdigitating loops leave virtually no gaps or clefts that might destabilize the pentamer structure (Fig. 6). Many of the contacts involve polar interactions. The tight interface extends only to the outer two-thirds of the pentamer, however, and there are no defined contacts between monomers near the pentamer base. This splaying of the subunits away from each other can be seen in Figure 6. The lack of interactions is reflected in the higher thermal factors for residues making up the inner face of the pentamer.

**CD-loops**

The CD-loop (residues 96–106) at the bottom of VP1 projects markedly from the molecule (Fig. 7a). Its structure is different in all six VP1 monomers, as shown in Figure 7b. The electron densities for these loops are among the weakest in the entire VP1 monomer. In some cases (monomers α, α' and α''), this disorder may be related to
the formation of a disulfide bond that involves Cys104 of different pentamers. Cys104 is located at the tip of the CD-loop, and because the loops of monomers α, α' and α'' approach each other in the capsid, we previously suggested that randomly selected disulfide bonds might be present between two of the three loops. Random covalent pairing can produce multiple loop conformations. The Ca–Ca distances between the closest Cys104 residues of monomers α, α' and α'' are within 7–10 Å, making disulfide bond formation possible. The CD-loops of monomers β and β' also approach each other, and their cysteines form a disulfide bond, which is now clearly defined by electron density. The CD-loops of β and β', although somewhat better defined by electron density than the remaining CD-loops, are still relatively mobile (see Fig. 2e). The CD-loops of monomers γ do not approach each other in the corresponding pentamer–pentamer contact and, with a Ca–Ca distance of 35 Å, the Cys104 residues cannot possibly form a disulfide bond. The cysteine of that monomer also points in a different direction (Fig. 7b). In regard to the disorder of the CD-loops, it is important to bear in mind the general mobility of the bottom part of the pentamers. There may be interactions between the CD-loops and the internal proteins VP2 and VP3, which do not obey icosahedral symmetry.

It is interesting to note that in murine polyomavirus, the structure of which has been determined at the significantly lower resolution of 3.65 Å, the CD-loops are much better defined. Several of the CD-loops in polyoma are even involved in the formation of short two-stranded antiparallel β-sheets [26]. In each of the CD-loops in polyoma, Cys114 (which corresponds to Cys104 in SV40) forms a disulfide bond with Cys19 at the N-terminal arm of a neighboring pentamer. This linkage prevents interpentamer disulfide bond formation between CD-loops and stabilizes the loop conformation. In SV40, Cys19 is replaced with proline. The Cys19–Cys114 disulfide bond in polyoma locks the N-terminal arm in a position next to the invading arm of another pentamer, and appears to make it difficult for the invading arm to withdraw. Although the conserved cysteine residues in the CD-loops of polyoma and SV40 are involved in very different disulfide pairings, these covalent interactions achieve the same goal: stabilization of the assembled particle. The fact that Cys104 is conserved in murine polyoma as well as in the human polyomaviruses, JC and BK [27], suggests that its role may indeed be of some importance.

The connector

The sequence Lys296-Asn297-Pro298-Tyr299-Pro300 forms a connector between the final (I) strand of the jelly roll and helix αC. It is at this point that the polypeptide chains of the six VP1 monomers α, α', α'', β, β' and γ enter non-equivalent regions in the virion. A close inspection of the connector residues reveals that they adopt three distinct conformations: the first is seen in α', β and β'; the second in α and α''; and the third, which is less well
ordered and not discussed further, in γ. A superposition of
the connector segments of monomers α, α', α'', β and β' is
shown in Figure 7c. The side chain of Asn297 forms two
well-defined hydrogen bonds to Tyr299-N and Tyr299-O
in monomers α', β and β', only one hydrogen bond (to
Tyr299-N) in α, and none in α''. The side-chain confor-
mation seen in α', β and β' may be a more favorable one,
as these residues have lower temperature factors and
make two hydrogen bonds. Thus, residues 296–300 may
be regarded as an extended hinged link, constrained by
the two proline residues. This sequence is completely
conserved in murine polyomavirus and in the human
polyomaviruses JC and BK.

The C-terminal helices
All VP1 molecules, with the exception of monomer γ,
have a helix, αC (residues 301–312), following the connec-
tor segment. This helix mediates contacts between pen-
tamers. The connector places the helix in a very exposed
position, in which it does not form any contacts with the
Figure 8

Interactions of the invading arm segment that precedes strand J with the target VP1 monomer. At the N-terminal end of the invading arm, the first interactions involve Val316, which points into a hydrophobic pocket around Tyr280, and Asp317, which forms two β-sheet-type main-chain hydrogen bonds with residues Asp198 and Asp200 on the EF-loop. An extensive hydrogen-bond network involves Gin319, which forms four hydrogen bonds to three different residues. The side chain of Gin319 points into a pocket where it interacts with the main chain of Asn52 and the side chain of Asp200. The absence of a side chain in Gin318, which precedes Gin319, is important for allowing Gin319 to approach the monomer closely. The putative cation-binding site is indicated by a green circle (Figure generated with GRASP [69]).

Monomers β and β' form a two-helix contact with hydrophobic interactions involving Ile301, Leu304, and Leu308 at positions a, d and a+7 of the heptad sequence. There are salt bridges between Glu307 and Arg311 at positions e and g: coiled-coils often have oppositely charged residues at these positions [29]. The crossing angle of the helices is ~60°, much larger than the ~20° angle in extended coiled-coil dimers [29,30]. Nevertheless, the side chains of the buried residues are packed in the standard conformation, and the helices fit snugly together. The total surface buried in the contact is 450 Å². Both helices and the preceding connector regions are characterized by low temperature factors. Furthermore, they have very similar conformations in both monomers β and β'.

Monomers α, α', and α'' form a three-helix contact. The interface created by the three helices is predominantly hydrophobic and involves Ile301, Leu304, Leu305 and Leu308. Again, there are a total of three salt bridges at the C termini of the helices, involving Asp307 and Arg312 of each helix in cyclic permutation. The total buried surface is 500 Å². The three helices diverge and are only loosely held together by hydrophobic interactions. The temperature factors of the α, α', and α'' helices are significantly higher than those of the β and β' helices, which form the two-helix contact. The three-helix contact is significantly asymmetric. It resembles a distorted two-helix contact to which one helix has been added, rather than a threefold bundle. The differences in connector conformations, described above, also contribute to this asymmetry.

The invading C-terminal arm

The C-terminal 49 residues (313–361) of one VP1 monomer lead away from it and invade another pentamer. The interactions made by the invading residues are the same for all six monomers: that is, the arm docks in a unique way. Residues 330–335 form strand J, which is part of the AJBIDG2 sheet. The preceding segment, residues 313–329, also forms a number of hydrogen bonds with the reference monomer. Side chain polar interactions between a monomer and its invading arm exclusively involve residues that are conserved in polyoma and SV40. The interactions of the invading arm with a VP1 monomer are shown in Figure 8. Residues 320–327 form a loop, here called the 'lid', which is stabilized by two internal main-chain hydrogen bonds, and its structure is conserved in polyoma. The lid points away from the invaded monomer, and it appears at first to interrupt the interactions between the C-terminal arm and its target. We believe that the lid helps to stabilize the proper conformation of the N-terminal 'clamp' (see below), because it lies just over the turn between strand A and helix 3A. Strands A/A' and J/J' participate in an extended β-sheet interaction, terminated by prolines at positions 20 and 341 at one end and by the clamp/lid interaction at the other. The extreme C-terminal part of the arm intersects with the clockwise neighbor of its principal target. There is a short β-sheet-like interaction with Lys298' at the end of strand I (Leu340-O...N-Lys298, Gly342-N...O-Lys298'). In addition, the side chain of Leu340 points into a hydrophobic pocket which also involves Val250' and Leu253' of the hairpin loop between strands G1' and G2'. The hairpin loop essentially plugs an opening at the rear,
causing the invading arm to turn at residue 341. These extensive interactions combine to bring Asp345 close to one of the ion-binding sites. In some of the monomers, residues 346-361 form a hairpin loop, which includes a short antiparallel β-ribbon (K10). In the other monomers, this segment is disordered.

The clamp
The clamp includes the N-terminal arm, strand A and helix 3A. This part of the structure integrates the invading strand J into the AIBIDG2 sheet. In addition to mainchain hydrogen bonds within the β-sheet (see Fig. 4), there are extensive hydrophobic interactions involving residues in strand A and helix 3A in the clamp, strand J, and strands B and C. These interactions appear to help fix the position of the invading strand J. There are also additional β-sheet interactions between A’ and J’ (Fig. 4).

It is extremely unlikely that the clamp segment retains the structure seen in the virus when the invading strand J is absent, that is, in an unassembled pentamer, because strand A and helix 3A depend upon J for stabilizing interactions. The lid (residues 320-327) of the invading C-terminal arm may play a role in determining the conformation of the clamp. The lid prevents an N-terminal extension of 3A and stabilizes the turn that precedes it. There are two glycine residues in this turn, which are conserved in polyoma.

The interface between strict and local pentamers
In the fully assembled virus shell, most pentamers contact each other only through the C-terminal arms as described above. There are two cases of a direct interpentamer contact involving the ‘body’ of the subunit. The first is an extensive and well-defined contact between strict and local pentamers, which involves residues with low temperature factors. The second occurs between the icoshedral dyad between two local pentamers, at the point where monomers γ approach each other very closely. This second contact, however, involves residues that are significantly disordered.

The contact between strict and local pentamers is centered at a pseudo-twofold symmetry axis (Fig. 9a), where the two pentamers exchange arms, so that the arm of one α monomer invades α’, and the arm of α” invades a second α (α’). The interface is substantial, with a buried surface area of 830 Å². About half of this area can be attributed to interactions involving residues 27–38 in α and α’. These are part of strand A and the following helix 3A. There are a number of intramonomer hydrophobic interactions between strand A and helix 3A: the side chains of Val27, Ile28, Val34 and Val37 hold together the strand-helix motif. This motif then packs against its counterpart in the neighboring pentamer, creating a hydrophobic core centered at the pseudo-twofold symmetry axis. There is also one hydrogen bond per monomer across the interface (between Lys38-N and Val27-O). Interactions involving

the invading arm account for the second half of the contact area. The invading arms of α and α” approach each other at Met324, and the two methionine side chains are within van der Waals distance. Residues 348–355, at the C termini of the invading arms, approach the 3A helices from the side opposite strand A (Fig. 9a). A ribbon drawing of the interface is shown in Figure 9b.
The contact between strict and local pentamers is unique in that it is formed by subunit faces approaching each other directly, with the αC helices flanking it on either side; all other interpentamer contacts in the virus are centered on emerging arms. The strict/local interpentamer contact may therefore have a role in defining particle curvature during the assembly process, because it involves the extended β-barrels of apposed subunits, rather than the more flexibly jointed arms.

**Cation-binding sites**

Calcium ions have been shown to facilitate *in vitro* assembly of murine polyoma VP1 pentamers in solution [31,32]. Our final electron-density map does not show any features that could be attributed to calcium ions. The crystals have been grown in high concentrations of sulfate, however, and the precipitation of calcium sulfate could have removed any calcium ions from the virus.

Following earlier work, we collected data from crystals soaked in 10 mM Gd(NO₃)₃ in order to locate potential calcium-binding sites as gadolinium can substitute for calcium. The \( (F_{\text{calc}} - F_{\text{nat}}) \exp(i\alpha_{\text{calc}}) \) difference map clearly shows two very strong positive density peaks per VP1 monomer at the 8g level. The centers of these two peaks, which are by far the highest in the entire difference map, are 9 Å apart. Figure 10a indicates the position of the two peaks in the structure. We refer to them as calcium sites 1 and 2.

Because the structure has been determined in the absence of calcium ions, the liganding side chains may not be in the same conformation as they would be if calcium were bound. The difference map does not show any side chain shifts, and we therefore cannot give precise ligand distances. We can, however, identify residues that are able to contact a calcium ion in one of their favored rotamer conformations, based upon a maximum distance of 4.0 Å (see Fig. 10b). A calcium ion at site 1 can be contacted by the carbonyl oxygen of Ser213' as well as by the side chains of Glu46, Glu48 and Glu216' in one VP1 pentamer and Glu330 from the invading strand J. Calcium at site 2 can be contacted by the carbonyl oxygen of Lys214' and by the side chains of Glu157', Glu160', Glu216' and Asp345. Thus, Glu216' could serve as a bridge between two ions.

Both sites link the invading arm with the pentamer core and obviously contribute to the stability of the particle by clamping together different pentamers. Site 2 was detected earlier by a similar soaking experiment [1]. Both sites have roughly similar density values in the difference map, suggesting comparable occupancy. Density for both sites is seen in all subunits, except in monomer γ, where the map does not show a peak for site 2. Indeed, the C terminus of the arm that invades the γ monomer, and in particular the calcium-liganding residue Asp345, does not assume an ordered structure. Similarly, we expect calcium only to bind to both sites 1 and 2 when the invading arm is present.

**Site for VP2/VP3**

The conical opening along the axis of the polyoma and SV40 VP1 pentamers (Fig. 6) shows electron-density features that may be attributed to the internal proteins VP2 and VP3 [1,33]. A hydrophobic segment of about 40 residues near the common C terminus of polyoma VP2 and VP3 is sufficient to bind to pentamers of polyoma VP1 [34]. The same is probably true for SV40. The residues lining the conical opening of VP1 are also largely hydrophobic. They lie on strands D, E, F, G1 and G2 as well as on loop DE. The interior surface of the conical opening is relatively flat, and no large side chains protrude from the protein surface. The narrowest part is formed by Pro231, at roughly two-thirds of the pentamer height. This residue is conserved in polyoma. It is unlikely that the internal proteins extend through the cone beyond this point, because the opening is only about 10 Å wide.

**Discussion**

**Virus assembly**

*An assembly model*

In our analysis of the refined SV40 model, we have described in detail the regions that play a role in assembly and capsid stabilization. These include the C-terminal helices, the invading arms with the lid, the clamps, the CD-loops, the disulfide bonds, the interface between strict and local pentamers, and the cation-binding sites. With the exception of the disulfide bonds, each of these structural features are present in both SV40 and polyoma, and the assembly process is likely to be very similar. We therefore consider data for both viruses in the following discussion.

Recombinant VP1 pentamers are capable of assembling into empty, capsid-like structures in solution [31,32]. These structures are not very homogeneous, but they show that the basic information for forming the virus shell is contained in VP1. Calcium ions, pH, and ionic strength influence assembly [32]. VP1 pentamers can also form two smaller capsid-like structures *in vitro*: a 12-pentamer icosahedron (12-ICOSA) and a 24-pentamer octahedron (24-OCTA) [32], as well as tube-like assemblies [35,36]. Although of no direct biological relevance, these incorrectly assembled structures can be viewed as products of an assembly pathway in which, at an intermediate stage, something has gone wrong. In projection, the VP1 pentamers have a distinct pentagonal shape, characterized by edges and faces. The image reconstructions of virions and of certain incorrectly assembled structures are detailed enough to determine the orientation of the pentamers on the basis of their shape [32]; by comparison with the virions, interactions in the aberrant assemblies can be
assigned. The VP1 pentagons can make face-to-face contacts, corresponding to the pseudo-twofold interaction seen between strict and local pentamers in the virion (marked by the asterisk in Fig. 1b), or two kinds of vertex-to-vertex contacts, corresponding to the β-β' and γ-γ interactions in the virion. The face-to-face contact seems to be a preferred interaction.

How can we picture the assembly process? An analysis of T-3 RNA virus assembly [37] demonstrated three important properties of that pathway. First, a defined, small oligomer (in the case of many T-3 structures, a dimer) constitutes the 'assembly unit'. Second, a specific nucleating structure initiates virion formation, which then proceeds by addition of assembly units to a growing shell.
rather than by coalescence of larger, stable intermediates. Third, the conformation and bonding properties of an assembly unit are determined uniquely and correctly by the characteristics of the site on the growing shell to which they add.

We believe that SV40 and polyoma capsids are likely to assemble by a mechanism with the same three properties. The assembly unit is, of course, the pentamer. We suggest that capsid formation might then occur as follows. The first step in our model is the encounter of two pentamers, P1 and P2 (Fig. 11). A pairwise interaction is much more likely than a three-pentamer encounter. We expect the C-terminal arms of the free pentamers to be disordered. Because of the predominance of face-to-face contacts in observed assemblies, we postulate that the most stable initial interaction involves interchange of arms in the pattern characteristic of the strict/local pentamer contact. We note that the other possible pairwise interchange, a vertex-to-vertex contact as seen in the β-β' interaction, lacks the additional interface shown in Figure 9. It is likely that neither pairwise interaction is very stable, and that only the face-to-face pair can recruit further pentamers rapidly enough to proceed to more stable intermediates and ultimately to capsids. Close contacts between pentamers also appear to be necessary to drive the assembly towards a curved structure and to prevent the formation of an essentially two-dimensional lattice, such as is seen in the tube-like structures [35,36].

Addition of a third pentamer, P3, to the P1-P2 pair can involve a similar face-to-face interaction and pairwise arm interchange (Fig. 11). A fully symmetric trimer with three identical face-to-face contacts can, however, form only by imposing a dramatic curvature. The smaller capsid structures observed by Salunke et al. [32] suggest that such a symmetric trimer of pentamers can occur, and that it is probably an assembly intermediate in the 12-ICOSA and 24-OCTA capsids (Fig. 11).

For proper formation of the virus capsid with its larger radius and smaller curvature, the trimer of pentamers must necessarily remain asymmetric (Fig. 11). Addition of three more pentamers to such a trimer can, however, create a 5-fold symmetric structure with the maximum number of tight interfaces possible. This structure corresponds to a strict pentamer surrounded by five local ones in the virus capsid, and this intermediate is already likely to have a relatively well-defined curvature. The symmetric six-pentamer structure can serve as a nucleus to select new additions in sensible ways. Pentamers cannot be added face-to-face without producing large holes, because they are not able to form face-to-face contacts with two pentamers of the nucleus simultaneously. They can, however, easily add by
forming two vertex-to-vertex contacts (Fig. 11). One of these will be a β-β' intersection; the other, γ-γ. The choice, which is dictated by the inherent chirality of the pentamer itself, determines the handedness of the underlying T=7 lattice. The important part about the addition of this pentamer is that its interactions are determined by the nature of the site at which it adds. Subsequent additions will also have a lowest energy configuration, leading to propagation and completion of the viral shell.

The C-terminal helices are short, and their stability appears to depend on their interactions in the assembled shell. Therefore, they are not likely to participate in the initial contacts between pentamers. Rather, we expect that formation of the C-terminal helices helps to rigidify an already assembled sector of the viral capsid. For example, once the postulated six-pentamer nucleus has formed, folding up of the helices and completion of the three-helix contacts may help to fix its curvature and thereby define the characteristics of the sites for addition of further pentamers.

Accuracy of assembly
What about mistakes (i.e., what happens when incorrect arm interactions occur?) The events postulated to lead to tubes and to the 12-ICOSA or 24-OCTA structures are examples of such mistakes. If one incorrect interaction occurs during icosahedral shell formation, assembly cannot proceed to successful completion. VP1 expressed in insect cells assembles in the nucleus into capsids that are much more homogeneous than those formed in vitro [38]. It is conceivable that cellular chaperone proteins play a role in assembly, perhaps by facilitating dissociation of ‘dead-end’ assembled structures and releasing assembly units to ‘try again’. This corrective function is indeed the way GroEL and related hsp60 proteins are thought to function in catalyzing protein folding [39].

The interactions that link together pentamers in SV40 and polyoma are locally just the same as those that complete a folded protein domain — addition of the last β-strands (J, A) to a barrel. Therefore, errors in SV40 assembly will produce intermediates that resemble ‘misfolded’ proteins. There will be exposed arms and unsatisfied sites for their attachment. Thus, it is plausible that mechanisms similar to those that operate in protein folding may permit escape from kinetic dead ends. There is also good evidence that polyoma VP2 or VP3 binds to VP1 pentamers in the cytosol, and that these complexes are transferred to the nucleus [34,40–42], where they assemble around the minichromosome. It has been shown that VP1 has affinity for DNA [43,44], but that the presence of viral DNA inhibits assembly of VP1 [43]. One can therefore imagine that the internal proteins VP2 and VP3 and the DNA influence the accuracy of assembly to some extent, even if they are not required for it.

Role of phosphorylation
Polyoma VP1 is phosphorylated after translation in infected cells; this modification probably occurs prior to assembly of the particle [45,46]. There is evidence that links polyoma mutants defective in encapsidation and growth with underphosphorylation of VP1 [46]. The residues that are being phosphorylated have not yet been identified unambiguously. A possible candidate for phosphorylation is Thr113 (Thr103 in SV40) (T. Benjamin, personal communication). This residue lies at the tip of the exposed CD-loop, which reaches into the interior of the shell (see Fig. 3b). The threonine is located within a stretch of five residues that are strictly conserved in polyoma, SV40, and the human JC and BK viruses. Because only a small fraction of VP1 in mature virions is phosphorylated [47], partial phosphorylation of the CD-loop could help explain its apparent disorder, which is indicated by poor electron density.

Viral entry
Non-enveloped viruses like SV40 must have some mechanism for entering the cytoplasm after binding to their receptors on the extra-cytoplasmic side of the cell membrane. In the case of SV40 and polyoma, mutational studies implicate VP2 in this process and, in particular, suggest a role for the N-terminal myristyl group [48,49]. Members of several groups of non-enveloped animal viruses contain myristylated structural proteins, which are buried in some way in the mature virion [50–53]. In all cases, the polyepitides that bear this modification are thought to have a role in penetration, and in some cases they are known to become exposed at a step during entry [54].

Picornaviruses undergo irreversible conformational changes upon interaction with cell-surface receptors [52]. In the case of poliovirus, this structural change involves particle expansion and exposure of the normally internal, myristylated VP4 as well as exposure of an internal part of VP1 [54]. Other picornaviruses appear to come apart upon receptor binding [55], thereby also exposing internal components. Because mutations in polyoma and SV40 VP2 appear to inhibit penetration, and because the myristyl group may help target VP2 to a membrane bilayer, it seems reasonable to think that VP2 must emerge at some stage to interact directly with the membrane of the cell being infected.

The tightly interlocked structure of the pentamer argues against a conformational change that would allow VP2 to exit along the 5-fold axis. Consistent with this view is our recent finding that the structure of the isolated, recombiant polyoma VP1 pentamer is essentially identical to its structure in the virion (TS and SCH, unpublished data). We therefore believe that expansion or dissociation would be required for exposure of VP2 in SV40 or polyoma. The structure does not immediately suggest a mechanism, however. We know from the complexes of polyoma with
receptor fragments, described in an accompanying paper [26], that polyoma does not expand or dissociate simply in response to receptor binding. Moreover, the peripheral location of the receptor site is not consistent with coupling to a profound conformational transition. Morphological studies of SV40 entry suggest that the virus is taken up by micropinocytosis and that virions then accumulate in structures associated with the endoplasmic reticulum (ER) [56]. If the pathway revealed by this approach is indeed the infectious route, dissociation must occur subsequent to ER accumulation. It is possible that interaction with a protein in the ER or in the space between the two membranes of the nuclear envelope triggers such a step.

It has been suggested that chelation of calcium ions and reduction of disulfide bonds are key events leading to uncoating of polyoma and SV40 [38,57]. Certainly these steps are important for dissociating the particle in vitro [58-60], and the structure suggests that they would also be essential in vivo. The calcium ion concentration and redox environment in the membrane-bound structures that receive the engulfed virions are not known, however. Abstraction of bound calcium does appear to play a role in uncoating of T-3 plant and insect viruses. In TBSV, for example, two calcium-binding sites are located at each intersubunit contact [61]. When calcium is removed and the pH is raised above 7.0, the interfaces are destabilized by charge-charge repulsion, causing the particle to expand. The SV40 calcium-binding sites described here appear to have similar characteristics, in that the side chains binding the cations belong to different pentamers. It is therefore unlikely that an isolated pentamer can bind calcium. Indeed, we believe that calcium binding and disulfide formation probably occur after the particle has assembled or even after its release. There is no strong evidence that either of these would occur in the nucleus, where SV40 and polyoma assemble. Calcium ions do influence the in vitro assembly of polyoma capsids, and in their absence high salt concentrations are necessary to drive capsid formation [31,32].

**Biological implications**

The structure of SV40, refined at 3.1 Å, has allowed us to analyze in detail the contacts between VP1 pentamers that stabilize the particle and direct its assembly. The way in which the C-terminal arms of VP1 tie the capsid together explains how a pentamer can interact specifically with six neighbors. Assembly must proceed by a mechanism that leads to correct choice of the type of interactions that each subunit makes. Based on our analysis of the SV40 structure, we propose a model for the assembly pathway with three essential characteristics: first, a defined assembly unit (the pentamer); second, a nucleating structure (the cluster of five pentamers around one); and third, growth by addition of individual assembly units, whose interactions are determined by the characteristics of the site on the growing shell to which they add.

Most of the interpentamer contacts involve folding the arm of a subunit in one pentamer into the β-barrel domain of a subunit in another pentamer, and we suggest that error-correction mechanisms, similar to those known to accelerate protein folding in vivo by alleviating kinetic dead ends, may further contribute to accurate assembly of virus particles. Bound divalent cations and disulfide bonds also stabilize the SV40 virion, probably by fixing the fully assembled structure. The stabilizing interactions are likely to be removed during entry. To understand steps of viral uncoating subsequent to receptor binding and uptake by micropinocytosis — in particular to work out the role of VP2 — further studies of the entry pathway are needed.

**Materials and methods**

**Crystalization and data collection**

Crystals of SV40 were grown from ammonium sulfate using the sitting drop method as described earlier [62]. They belong to space group I23 with a=56.8 Å and contain two complete virions in the unit cell. The crystallographic asymmetric unit contains 30 VP1 monomers; a pentavalent 'strict' pentamer and five hexavalent 'local' pentamers. X-ray data were collected at the Cornell High Energy Synchrotron Source (CHESS) F1 beamline using the oscillation method and a wavelength of 0.91 Å. The characteristics of this beamline — very high brilliance and low divergence — were essential for data collection. In total, 46 crystals of a thickness of at least 500 μm were used for the collection of the 3.1 Å data set. The crystals were randomly orientated in the beam, and typical exposure times were 20–40 s. The crystals were cooled to −15°C during data collection. Only one diffraction image per crystal volume was recorded. Typically, 2–3 exposures could be obtained per crystal using a 0.15 mm collimator, which is about the largest diameter that still allows adequate spot separation for the 3.1 Å data. The oscillation angle was 0.2°. Data were recorded on 200 × 250 mm Fuji Imaging Plates (Fuji Inc., Japan). The plate sizes permit a long crystal-to-detector distance of 400 mm, and we were able to take full advantage of this property, because a flight tube constructed by CHESS allowed the crystal-to-plater distance to be changed without altering the focusing point of the beam. The detector distance was set such that the short edge of the plate corresponded to 3.55 Å resolution. The loss of the rather sparse data close to the spindle was judged less important when compared with the improvement in spot resolution at the somewhat longer crystal-to-plater distance. Three adaptations were made to the setup in order to reduce the background radiation. (i) Extra lead shielding was installed in the CHESS F1 hutch around the sitting devices at the distant end of the flight tube; a lead collar was fitted around the shutter housing; and only one image plate left inside the hutch during data collection. (ii) A sliding cap was made for the X-ray collimator and a very small (<0.5 mm) beamstop was made on a special holder. These two together reduce the air path from which is normally 50–60 mm to just 4 mm. (iii) The Fuji BAS2000 scanner was recalibrated with a uniform flood field just prior to data collection.

The images were indexed and integrated using DENZO (Z Otwonowski, Yale University). Because different crystals diffracted to somewhat different resolution ranges, each integrated image was filtered before scaling. Each integrated file was scanned and a histogram calculated of <f²>/ <f> versus resolution. An individual resolution limit was applied to each frame according to a shell cutoff requirement of 1.5 (<f>).</p>

Scanning and merging were carried out using ROTAVATA/AGROVATA.
Table 1

Data statistics.

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>$R_{exp}$ (%)</th>
<th>$&lt;d_{ref}&gt;$</th>
<th>Unique reflections</th>
<th>Completeness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12–5.8</td>
<td>5.8</td>
<td>10.6</td>
<td>60405</td>
<td>85.9</td>
</tr>
<tr>
<td>5.8–4.7</td>
<td>6.4</td>
<td>9.0</td>
<td>59136</td>
<td>85.4</td>
</tr>
<tr>
<td>4.7–4.1</td>
<td>8.1</td>
<td>7.6</td>
<td>63557</td>
<td>85.3</td>
</tr>
<tr>
<td>4.1–3.8</td>
<td>14.1</td>
<td>4.7</td>
<td>48004</td>
<td>84.2</td>
</tr>
<tr>
<td>3.8–3.5</td>
<td>24.7</td>
<td>2.6</td>
<td>64337</td>
<td>82.2</td>
</tr>
<tr>
<td>3.5–3.3</td>
<td>30.6</td>
<td>2.2</td>
<td>52495</td>
<td>76.0</td>
</tr>
<tr>
<td>3.3–3.1</td>
<td>44.1</td>
<td>1.6</td>
<td>54413</td>
<td>62.6</td>
</tr>
<tr>
<td>Total</td>
<td>9.7</td>
<td>6.5</td>
<td>402347</td>
<td>80.2</td>
</tr>
</tbody>
</table>

*R$_{exp}$=$\sum_{h}$$|I_{h}|$/$\sum_{h}$$<I_{h}>$, where $h$ are unique reflection indices, $I_{h}$ are intensities of redundant, symmetry-related reflections of index $h$ and $I_{e}$ is the mean intensity for reflections of index $h$.

(CCP4 program suite [63]). Post-refinement was also carried out using the CCP4 program package. For the final data set, 125×0.2° images were scaled with a reference data set also collected on Fuji image plates. Subsequent scaling and data merging analysis revealed no systematic bias between fully and partially recorded measurements. The final data set comprised all fully recorded measurements (after normal rejection tests) and scaled-up partial measurements with a fraction recorded >0.5. Finally, intensity data from AGROVATA were run through TRUNCATE [64] to apply the usual Bayesian remapping of the weakest amplitudes.

The final data set is 80.2% complete and consists of 402,347 unique reflections between 12 Å and 3.1 Å. Of these, 8008 reflections were set aside as a 'free' data set [65], and these reflections were excluded from every calculation performed during the structure determination described below. The remaining reflections were used without applying any cutoff. The final data statistics are given in Table 1. The $I_{rel}$ ratio is close to 1.0 at the outer resolution edge of the data, accounting for the high internal R-factors and low accuracy of the measurements beyond 3.5 Å.

Phase extension and refinement

The initial R-factor, using the new data set and the unreified model of SV40 [1], was 51.5% (12–3.5 Å) after rigid-body refinement. We then carried out energy minimization with X-PLOR [5], using a refinement procedure similar to the one used in the structure determination of polyoma [26]. Only 5-fold strict NCS was imposed as a constraint in X-PLOR. The use of the strict NCS proved to be very powerful for constraining the refinement. Thus, the R-factors for the 'working' set and the free set behaved in a very similar way, with the $R_{free}$ being only 0.005–0.015 larger than $R_{exp}$. A result of the 5-fold NCS is that the $R_{free}$ is no longer completely independent, and we quote only working R-factors below.

Initially, all atoms were given temperature factors of B=50 Å², a value derived from a Wilson plot. Positional refinement (200 cycles, resolution range 12–3.5 Å) reduced the R-factor to 37.2%. A subsequent temperature factor refinement followed by an additional 100 cycles of positional refinement brought the R-factor down to 30.4%. The refinement appeared to stall at that stage. The free correlation coefficients between the observed amplitudes and the ones calculated from the model were 0.8 at 10 Å resolution, and 0.6 at 5 Å resolution. These values dropped dramatically, however, at higher resolution. In order to obtain an electron-density map essentially free of model bias we

Table 2

Course of refinement.*

<table>
<thead>
<tr>
<th>Round</th>
<th>Resolution range (Å)</th>
<th>RIGID$^a$</th>
<th>MIN$^b$</th>
<th>BREF$^c$ (cycles)</th>
<th>MIN</th>
<th>R-factor (start)</th>
<th>R-factor (final)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12–6.0</td>
<td>60</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.544</td>
<td>0.458</td>
</tr>
<tr>
<td>2</td>
<td>12–3.5</td>
<td>-</td>
<td>227</td>
<td>15</td>
<td>200</td>
<td>0.515</td>
<td>0.304</td>
</tr>
<tr>
<td>3</td>
<td>12–3.1</td>
<td>-</td>
<td>300</td>
<td>15</td>
<td>250</td>
<td>0.360</td>
<td>0.273</td>
</tr>
<tr>
<td>4</td>
<td>12–3.1</td>
<td>-</td>
<td>90</td>
<td>15</td>
<td>75</td>
<td>0.362</td>
<td>0.260</td>
</tr>
<tr>
<td>5</td>
<td>12–3.1</td>
<td>-</td>
<td>100</td>
<td>15</td>
<td>100</td>
<td>0.283</td>
<td>0.254</td>
</tr>
</tbody>
</table>

*The refinement was carried out with X-PLOR [5]. $^a$RIGID: rigid-body refinement (using the correlation coefficient as a target). Each VP1 monomer was treated as an entire subunit. $^b$MIN: conjugate-gradient minimization with CHARMM non-bonded potential. The value for the tolerance ranged from 0.1 to 0.05. $^c$BREF: individual B-factor refinement. The standard deviations between B-factors of bonded atoms and B-factors of atoms connected by an angle were restrained to 1.5 Å and 2.0 Å, respectively, for main-chain atoms and to 2.0 Å and 2.5 Å, respectively, for side-chain atoms.

Figure 12

Accuracy of the phase extension as monitored by the resolution-dependence of the free correlation coefficient. The final correlations for each new added shell are shown, both after five cycles of averaging at that resolution (solid line) and after completion of the phase extension at 3.3 Å (dotes). The loss of correlation at ~3.3 Å is obvious. The curves show that the correlation coefficient within a shell still improves noticeably when more shells are added. The correlation coefficient for the complete resolution range is also shown (broken line).
decided to carry out a phase extension procedure starting at 5.0 Å by making use of the NCS of the particle. At this resolution, the free correlation coefficient of 0.6 indicated a still reasonable agreement between model and observed data.

The starting phases from 12–5.0 Å were extended by 5-fold NCS averaging with the program package RAVE [4]. First, an envelope covering all model atoms present in the crystallographic asymmetric unit was produced with ENVAT (D Madden, Harvard University), using a radius of 5.0 Å. The envelope also covered the shallow pocket in the interior of a VP1 pentamer in order to include potential features from internal proteins. For averaging, we used a redimensioned version of AVE (VIRUS_AVE) installed on a DEC AXP 3000 running OS/1 (Digital Equipment, USA). The CCP4 program package [63] was used for structure factor and map calculations. All calculations were carried out in space group I222 (which is a subgroup of I23), with the exception of the structure factor calculation. This was performed in space group P21/21a using a map shifted in origin by (¼, ¼, ¼). That is, the map was first shifted to the new origin; the structure factors were calculated in the new space group; the phases of appropriate reflections were then corrected by ±180°. Five cycles of averaging were carried out at each resolution step; the average phase change in the last cycle was on the order of 4–5° and indicated reasonable convergence. The cycles were then followed by a resolution extension where the step size was one half of a reciprocal lattice point. Calculated phases were used as a starting point in the added resolution shell. A grid size of 432 x 432 x 432 was chosen for the entire phase extension from 5.0 Å to 3.35 Å. This grid is rather coarse at higher resolution; it was retained nevertheless to avoid a dramatic increase in computation time due to the limited workstation core memory. For this extension from 3.6 Å on, the data set was 'sharpened' by rescaling it with a temperature factor of 40 Å², which essentially resulted in a flat intensity distribution beyond 4.0 Å. The sharpening understandably led to a slight decrease of the free correlation coefficients but produced noticeably cleaner and better maps. Correlation coefficients using the free data set were calculated at each step to monitor the reliability of the averaging process. Typical starting values for free correlation coefficients in new resolution shells were 0.2–0.3; these numbers increased during averaging to the final values shown in Figure 12. At 3.3 Å, the free correlation coefficient did not improve noticeably, and the process was halted. The refined phases between 12 Å and 3.35 Å were then used to calculate a (2Fₒ-Fᵣ) map for manual rebuilding. In total, 350 cycles of averaging were carried out, with one cycle of averaging using about 1–1.5 h of CPU time.

The final map at 3.35 Å was of exceptional clarity, showing most side chains and the orientation of many main-chain peptide bonds. The SV40 model was completely rebuilt using O [7]. The corrected model was then again subjected to 200 cycles of energy minimization in X-PLOR, using all available data between 12 Å and 3.1 Å, which reduced the R-factor to 23.7%. Averaging (six cycles) was then carried out at 3.1 Å, using a finer grid of 12 x 12 x 12. After two more rounds of rebuilding and energy minimization, the R-factor converged at 25.4%. An overview of the refinement is given in Table 2. The final model consists of 15989 non-hydrogen protein atoms in the isocahedral asymmetric unit. It does not contain water molecules. VP1 monomers α, β, and γ contain amino acids 14–361, monomers α and α'' contain residues 14–355, and monomer β contains residues 14–343. Surface calculations were carried out with SURFACE (CCP4 program package) [63] using a probe radius of 1.4 Å. Coordinates have been deposited in the Brookhaven Protein Data Bank (accession number 1SVA).

Acknowledgements
We thank members of the Harrison and Wiley research groups for discussion and comments on the manuscript. We acknowledge CHRESS for access to the F1 beamline, and we thank the staff of CHRESS and MacCHESS for help. This work was supported by NIH grant CA 52022 (to SCH) and by the Howard Hughes Medical Institute. TS acknowledges support from an EMBO postdoctoral fellowship during part of this work. SCH is an investigator in the Howard Hughes Medical Institute.

References