The EMBO Journal Vol.16 No.16 pp.5139–5148, 1997

High-resolution structure of a polyomavirus VP1–oligosaccharide complex: implications for assembly and receptor binding

Thilo Stehle and Stephen C.Harrison[1]

Howard Hughes Medical Institute and Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Ave, Cambridge, MA 02138, USA

[1]Corresponding author
e-mail: harrison@crystal.harvard.edu

The crystal structure of a recombinant polyomavirus VP1 pentamer (residues 32–320) in complex with a branched disialylated hexasaccharide receptor fragment has been determined at 1.9 Å resolution. The result extends our understanding of oligosaccharide receptor recognition. It also suggests a mechanism for enhancing the fidelity of virus assembly. We have previously described the structure of the complete polyomavirus particle complexed with this receptor fragment at 3.65 Å. The model presented here offers a much more refined view of the interactions that determine carbohydrate recognition and allows us to assign additional specific contacts, in particular those involving the (α2,6)-linked, branching sialic acid. The structure of the unliganded VP1 pentamer, determined independently, shows that the oligosaccharide fits into a preformed groove and induces no measurable structural rearrangements. A comparison with assembled VP1 in the virus capsid reveals a rearrangement of residues 32–45 at the base of the pentamer. This segment may help prevent the formation of incorrectly assembled particles by reducing the likelihood that the C-terminal arm will fold back into its pentamer of origin.

Keywords: carbohydrate recognition/polyomavirus/sialic acid/viral assembly/virus–receptor interaction

Introduction

Murine polyomavirus ('polyoma') is a non-enveloped DNA tumor virus. The particle has a diameter of ~500 Å. Its shell contains 360 copies of the major capsid protein VP1 (45 kDa) arranged in pentamers on a $T = 7$d icosahedral lattice. The shell encloses the minor capsid proteins, VP2 and VP3, and the DNA genome (Tooze, 1981; Eckhart, 1990). Crystallographic analyses of the closely related and structurally very similar simian virus 40 (SV40) (Liddington *et al.*, 1991; Stehle *et al.*, 1996) and of polyoma itself (Stehle *et al.*, 1994; Stehle and Harrison, 1996) have revealed the three-dimensional structure of VP1 and the architecture of the virion shell. The VP1 polypeptide chain folds into a β-barrel with 'jelly-roll' topology; there are extensive contacts between VP1 monomers in a pentamer. The most striking feature of the capsid structure of polyoma and SV40 is the way the individual pentamers are tied together by the C-terminal arms of the VP1 monomers. The last 63 residues of VP1 emerge from each monomer and 'invade' another pentamer, where they form a β-strand that augments a sheet in the target subunit (Liddington *et al.*, 1991; Stehle *et al.*, 1996).

Polyoma recognizes oligosaccharides that terminate in (α2,3)-linked α-5-*N*-acetyl neuraminic acid (sialic acid; NeuNAc) on the surface of susceptible cells. While all strains bind to straight-chain receptors [e.g. NeuNAc–(α2,3)–Gal–(β1,3)–GalNAc], some strains also recognize branched oligosaccharides that carry a second, (α2,6)-linked sialic acid (e.g. NeuNAc–(α2,3)–Gal–(β1,3)–[(α2,6)–NeuNAc]–GalNAc) (Fried *et al.*, 1981; Cahan *et al.*, 1983). The ability to bind to these branched-chain receptors correlates with a single amino acid substitution at position 91 on the outer surface of VP1 and with a significantly decreased tumorigenicity. Low-tumorigenicity strains bear a glycine at position 91 and bind both receptor types; high-tumorigenicity strains bear a glutamic acid and primarily recognize the straight-chain receptor fragment (Freund *et al.*, 1991). The work described below was carried out using the low-tumorigenicity strain P16, which has a glycine at position 91. The structure of the complete virus particle in complex with model compounds for both receptor types, determined at 3.65 Å, has shown that these receptor fragments bind in a shallow groove formed by several loops at the outer margin of VP1 (Stehle *et al.*, 1994; Stehle and Harrison, 1996). In the more highly tumorigenic strains, discrimination between the branched and the straight-chain receptors comes from the Glu91 side chain, which interferes with the carboxylate of the branching sialic acid (Stehle and Harrison, 1996).

We present here the high-resolution structure of a recombinant VP1 pentamer complexed with a branched-chain receptor fragment. VP1 pentamers were first produced in *Escherichia coli* over a decade ago with the intention of crystallizing them. Although these pentamers failed to crystallize, they formed capsid-like structures (Salunke *et al.*, 1986, 1989). Since the C-terminal arms play an important role in capsid formation, we used a VP1 pentamer fragment that lacks the C-terminal residues 321–383 to prevent it from assembling. Crystals could only be obtained after cleavage of the N-terminal 31 residues with trypsin. The truncations at both termini are far from the carbohydrate binding site, and they do not affect the structure of the β-barrel. The one difference between the assembled and unassembled pentamer is near the adjacent N- and C-termini of the truncated protein. We suggest that this rearrangement may have a role in regulating assembly.

Results and discussion

Overall structure and model accuracy

The model for the complex comprises residues 32–316 for monomers 1, 3, 4 and 5 and residues 32–320 for

monomer 2, as well as five oligosaccharide fragments and 2127 water molecules. The structure of the recombinant VP1 pentamer, shown in Figure 1A, is very similar to that of the VP1 pentamer in the virus particle (Stehle *et al.*, 1994; Stehle and Harrison, 1996) but far more accurate. Each subunit has two antiparallel β-sheets with jelly-roll topology, and some of the loops that connect the β-strands are extensive and contain additional secondary structure elements (a small three-stranded antiparallel β-sheet, one α-helix and four 3_{10}-helices). An assignment of the secondary structure is given in Figure 1B. The α-helix (αD) and two of the four 3_{10}-helices (3A′ and 3A″) were not described in the virus structure at 3.65 Å resolution (Stehle and Harrison, 1996). The interface between two monomers is substantial. In particular helix 3B and the DE-loop insert deeply into a cleft in the clockwise adjacent VP1 monomer, and part of strand G ('G1') augments its CHEF sheet (Figure 1A).

The only significant differences between the structure of the recombinant VP1 pentamer and its structure in the assembled virus particle are the lack of the invading arms and a rearrangement of residues 32–45 at the N-termini. These differences and their possible implications are discussed in a later section. The remaining 271 Cα-atoms residues can be superimposed with root mean square (r.m.s.) deviations of 1.0 Å for individual monomers and 1.2 Å for complete pentamers.

Only a few regions of the protein are not well ordered; these include Gly32 and Gly33 at the N-terminus, the CD-loops (residues 105–120) between β-strands C and D at the base of the molecules, and the C-terminal residues 317–320, which are ordered only in monomer 2 because they participate in a crystal contact. The structure of the complete virus particle has shown earlier that the CD-loops do not assume a unique conformation. In the virus, their conformation is somewhat stabilized by a disulfide bond between Cys114 at the tip of the loop and Cys19 at the N-terminus of a clockwise neighbor (Stehle and Harrison, 1996). Since the structure of the recombinant fragment lacks residues 1–31 at the N termini, these disulfide bonds are not present in our structure. All residues fall in favorable regions of a Ramachandran diagram (Ramachandran and Sasisekharan, 1968). The coordinate error for the recombinant VP1 pentamer is about 0.15 Å, as estimated from a Luzzati plot (Luzzati, 1952).

The ligand used in this structure analysis is a Y-shaped hexasaccharide that contains two terminal sialic acids: NeuNAc-1–(α2,3)–Gal–(β1,3)–[(α2,6)-NeuNAc-2]–GlcNAc–(β1,3)–Gal–(β1,4)–Glc. The NeuNAc-1–(α2,3)–Gal–(β1,3)–GlcNAc part is well ordered in all five monomers. By contrast, the (α2,6)-linked sialic acid is better ordered in monomer 4 than in the others, probably because of hydrogen bonds between Tyr294-OH and Asn149-O in a crystallographically related molecule and the carboxylate and N5 respectively, of the sialic acid. Though less well ordered, the (α2,6)-linked sialic acids, and in particular the glycerol side chains, appear (according to the final electron density map) to have a similar position in the other four monomers. The following description will therefore be based on the complex seen in monomer 4. The initial difference Fourier map for the oligosaccharide in monomer 4 is shown in Figure 2. The (β1,3)–Gal–(β1,4)–Glc part (the 'stem') of the oligosaccharide is not ordered in any of the monomers and has not been included in the model.

Carbohydrate binding site

The crystal structure of a complete polyomavirus particle complexed with the hexasaccharide used in this study was first determined at 3.65 Å resolution (Stehle and Harrison, 1996). The description of the binding site and the key interactions listed in the earlier paper remain correct, but the high-resolution model offers a much more detailed and accurate view of the interactions that take place in the binding site (Figure 3). In particular, several additional hydrogen bonds between protein and ligand can be seen at high resolution, and also a substantial number of water-mediated hydrogen bonds. We also now see an extensive hydrogen bond network that is involved in recognizing the (α2,6)-linked NeuNAc-2. An overview of the interactions is given in Figure 3A.

The surface of VP1 is essentially a plateau with binding sites for the two branches of the oligosaccharide located at each end (Figure 3B). Each site is closed off at the far end by protein residues. NeuNAc-1 inserts into a groove formed by loops at each side, with the sialic acid atom O4 and the carboxylate group facing the protein and the glycerol chain projecting towards the solvent (Figure 3B). The *N*-acetyl group inserts most deeply into the groove, but the surface representation in Figure 3B shows that the far end of the groove is considerably larger than would be required to accommodate just the *N*-acetyl. The galactose is located at the entrance of the groove. Residues Arg77 and Gly78 form a ridge that separates NeuNAc-1 and Gal, creating pockets for NeuNAc-1 and Gal on either side of the ridge. NeuNAc-2 inserts into a shallow pocket at the other end of the VP1 surface. The surface area buried by each of the two sialic acids is very similar; the values are 175 $Å^2$ for NeuNAc-1 and 162 $Å^2$ for NeuNAc-2 [calculated with SURFACE (CCP4, 1994)]. The GlcNAc moiety is not contacted by protein. It is located above the flat surface that opens both toward the outside of the pentamer and toward the 5-fold symmetry axis (Figure 3B).

Site for NeuNAc-1–(α2,3)–Gal

The interaction between NeuNAc-1 and VP1 is centered at amino acids Arg77 and Gly78 (Figure 3C). These two residues form an L-shaped structure, similar to a bent arm, with which the NeuNAc-1 carboxylate and the Gal-O4 are hydrogen bonded. Thus, this two-residue segment primarily determines the specificity for NeuNAc–(α2,3)–Gal. The presence of glycine at position 78 appears to be critical. Its dihedral angles ($\phi = -54°$, $\psi = 123°$) are not exceptional, but any side chain, even the methyl group of an alanine, would sterically interfere with the galactose. The guanidinium of Arg77 is stacked against the aromatic ring of Tyr72, which in turn is hydrogen bonded to the N5 of NeuNAc-1. The Arg77 side chain is also hydrogen bonded to Gln59. There are several additional protein–oligosaccharide hydrogen bonds: O4 of NeuNAc-1 hydrogen bonds with His298-NE and Asn293-ND, and Gal-O6 hydrogen bonds with Asn93-ND. The side chain of His298 is stacked against the guanidinium group of Arg289. It is quite striking that, with the exception of Asn293, all residues that hydrogen bond with the NeuNAc-1–Gal moiety are located at the bottom of the groove and project

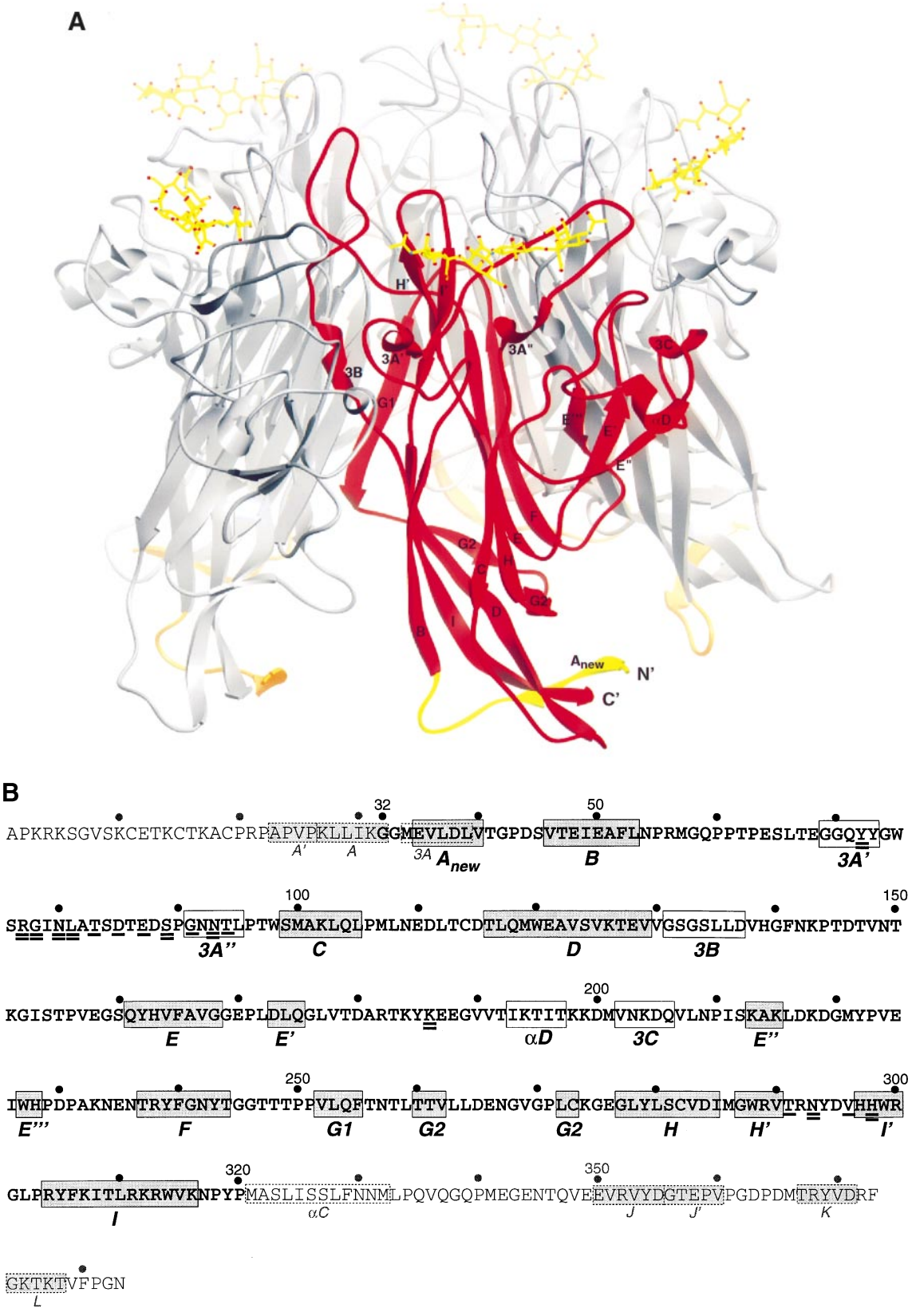

Fig. 1. Structure of the recombinant VP1 pentamer. (**A**) Ribbon drawing of the VP1 pentamer complexed with the disialylated oligosaccharide. One monomer is shown in red, the others in gray. The receptor fragments are shown as ball-and-stick models. The rearranged N-terminal segments are shown in yellow (red monomer) and orange. (**B**) Sequence of VP1 with secondary structure assignments based on hydrogen bond formation and main chain conformation. Residues 32–320, which correspond to the crystallized fragment, are shown in bold. The assignments for the remaining residues (shown in broken lines) are taken from the structure of the polyomavirus capsid (Stehle and Harrison, 1996). Residues that are within 4.5 Å of a carbohydrate atom are underlined; residues that hydrogen bond to the receptor fragment are underlined twice. Panel (A) was prepared with RIBBONS (Carson, 1987).

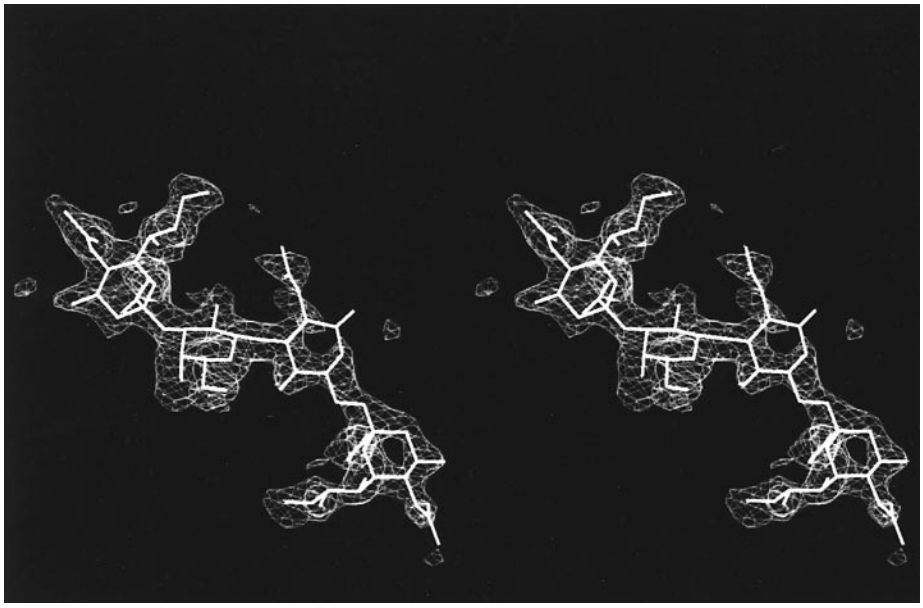

Fig. 2. Difference Fourier electron density, in stereo, for the oligosaccharide, calculated at 2.0 Å resolution and contoured at 2.5σ. Figure prepared with O (Jones *et al.*, 1991).

upward, forming a shelf on which the oligosaccharide rests (Figure 3D). These residues are very well ordered in the absence of the ligand. Asn293, at the tip of the HI-loop, projects downward and is less well ordered.

Several hydrophobic interactions are present in the binding site. The most striking such contacts are made by Thr291 and Val296: the two side chains face NeuNAc-1 from the rear and appear to wedge the carbohydrate ring into its pocket (Figure 3D). A mutation to Ala at position 296, which essentially eliminates the hydrophobic interaction of that residue with the carbohydrate, reduces the affinity of the virus for the oligosaccharide (T.Stehle and S.C.Harrison, unpublished results) but increases the pathogenicity of the virus dramatically (Bauer *et al.*, 1995). Two additional hydrophobic contacts are seen between Thr94-CG and Gal-C6, and between Gly91-CA and Gal-O6. Asp85# at the tip of the BC2-loop of the clockwise VP1 neighbor contacts the methyl group of the *N*-acetyl chain and closes the groove to the left (Figure 3B and C). *N*-glycolyl neuraminic acid, a sialic acid variant that occurs in murine cells (Kawano *et al.*, 1995), carries a CH_2–OH group instead of the CH_3 group at the end of this chain. This additional hydroxyl group would be in an ideal position to hydrogen bond with the side chains of Asp85# and Tyr72. It is therefore likely that polyoma will also bind, perhaps even with higher affinity, to oligosaccharides terminating in *N*-glycolyl neuraminic acids.

There are four internal hydrogen bonds in the oligosaccharide: within each sialic acid, between O8 and the carboxylate oxygen O1A; between the endocyclic oxygen of NeuNAc-1 and Gal-O2; between the endocyclic oxygen of Gal and GlcNAc-O4. Only three water molecules mediate additional hydrogen bonds between the sialic acid and VP1. Two of these water molecules are located at atom O4, the remaining one hydrogen bonds with O7 (Figure 3D).

Site for NeuNAc-2

A significant thermal-parameter gradient along the carbohydrate chain (NeuNAc-1: 26.0 $Å^2$, Gal: 31.9 $Å^2$, GlcNAc, 50.8 $Å^2$, NeuNAc-2: 55.9 $Å^2$) suggests that NeuNAc-2 is more loosely bound to the protein that the NeuNAc-1–(α2,3)–Gal moiety. This gradient may in part be attributed to the fact that the glycosidic (α2,6)-linkage can have multiple conformations. Nevertheless, there are specific interactions between NeuNAc-2 and the protein (Figure 3E). The glycerol side chain and the carboxylate insert into a shallow pocket that is formed by the side chains of Thr83, Glu87 and Lys186. Thr83 and Glu87 form hydrophobic contacts with the glycerol chain and the *N*-acetyl group at the far end of the binding site (Figure 3E). A hydrogen bond network at the glycerol side chain involves Asn80, Leu81 and Ser89. Lys186 forms a salt bridge to the sialic acid carboxylate. There are also hydrogen bonds linking Asn80, Ser89 and Lys186, and two water-mediated hydrogen bond bridges to NeuNAc-2.

No structural changes upon receptor binding

No dramatic structural changes occur upon binding of the receptor fragment. In fact, the side chains that contact the ligand are almost exactly in their correct position prior to complex formation, ready to receive the oligosaccharide. The Arg77 side chain rests against the aromatic ring of Tyr72, which probably helps stabilize its conformation, and the tip of the Arg side chain moves only 0.7 Å upon complex formation. Most of the carbohydrate hydroxyl groups that interact with the protein are replaced by water molecules in the unliganded structure (NeuNAc-1-O1A and -O4, Gal-O4 and -O6, NeuNAc-2-O8 and -O9), allowing the amino acid side chains to make very similar interactions in both cases. Thus, the receptor binds in a preformed groove, where it merely replaces water molecules. The binding site appears to be designed to primarily recognize the NeuNAc-1–(α2,3)–Gal moiety, and this

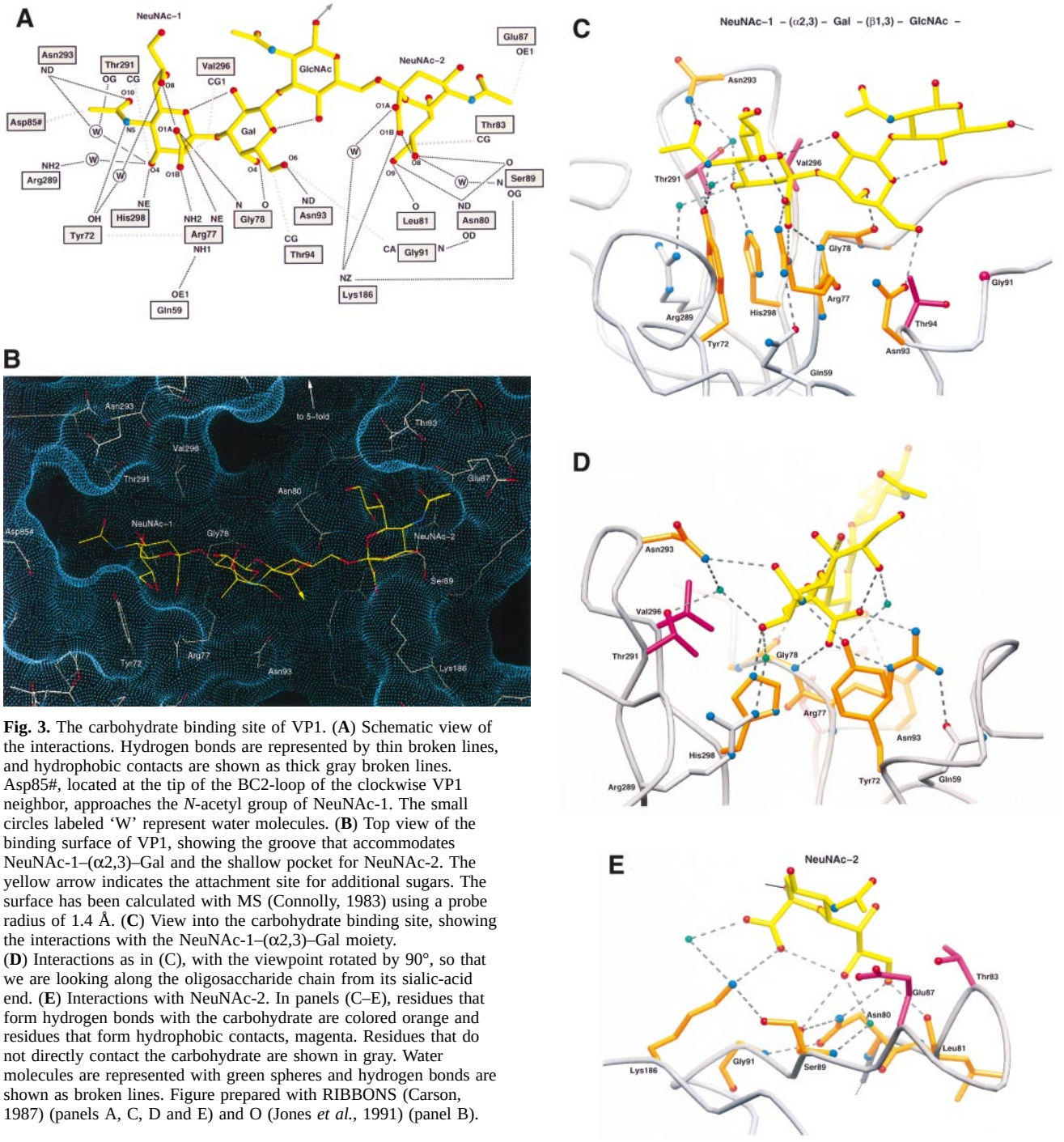

Fig. 3. The carbohydrate binding site of VP1. (**A**) Schematic view of the interactions. Hydrogen bonds are represented by thin broken lines, and hydrophobic contacts are shown as thick gray broken lines. Asp85#, located at the tip of the BC2-loop of the clockwise VP1 neighbor, approaches the *N*-acetyl group of NeuNAc-1. The small circles labeled 'W' represent water molecules. (**B**) Top view of the binding surface of VP1, showing the groove that accommodates NeuNAc-1–(α2,3)–Gal and the shallow pocket for NeuNAc-2. The yellow arrow indicates the attachment site for additional sugars. The surface has been calculated with MS (Connolly, 1983) using a probe radius of 1.4 Å. (**C**) View into the carbohydrate binding site, showing the interactions with the NeuNAc-1–(α2,3)–Gal moiety. (**D**) Interactions as in (C), with the viewpoint rotated by 90°, so that we are looking along the oligosaccharide chain from its sialic-acid end. (**E**) Interactions with NeuNAc-2. In panels (C–E), residues that form hydrogen bonds with the carbohydrate are colored orange and residues that form hydrophobic contacts, magenta. Residues that do not directly contact the carbohydrate are shown in gray. Water molecules are represented with green spheres and hydrogen bonds are shown as broken lines. Figure prepared with RIBBONS (Carson, 1987) (panels A, C, D and E) and O (Jones *et al.*, 1991) (panel B).

fragment probably has a very similar conformation in solution, with an internal hydrogen bond fixing the orientation between the two sugars.

Comparison with other sialyloligosaccharide complexes

The structures of several other protein–sialyloligosaccharide complexes have been reported at atomic resolution: (i) wheat germ agglutinin complexed with 3′sialyl lactose (Wright, 1990) and with a sialyloglycopeptide from glycophorin A (Wright and Jaeger, 1993); (ii) cholera toxin complexed with a ganglioside pentasaccharide (Merritt *et al.*, 1994); (iii) a legume lectin complexed with a biantennary glycopeptide (Bourne *et al.*, 1994); (iv) pertussis toxin in complex with a soluble oligosaccharide from transferrin (Stein *et al.*, 1994); (v) influenza hemagglutinin ligated with various sialyloligosaccharides (Sauter *et al.*, 1992; Eisen *et al.*, 1997).

The specific interactions between sialic acid and the protein are very different in each case, but the overall architecture of the binding sites and the way the sialic acid fits into the site seem to fall into two distinct classes. In the NeuNAc-1 site of polyoma and the wheat germ agglutinin–sialopeptide complex, the sialic acid inserts into a groove with residues on both sides of the groove facing the carbohydrate ring. The O4 atom is buried and

Table I. Dihedral angles[a]

	NeuNAc–(α2,3)–Gal (ϕ_1, ψ_1)	NeuNAc–(α2,6)–{GlcNAc/GalNAc/Gal} (ϕ_2, ψ_2, ω_2)	
Polyoma VP1	(–34, –5)	(–84, 162, 64)	{GlcNAc}
CWG[b] site A	(–46, 5)	(–62, –169, 78)	{GalNAc}
WGA[c] site A	(–69, –12)		
site B	(–67, –1)		
Cholera toxin[d]	(–169, –31)		
HA-23 SL[e]	(–62, –12)		
HA-26 SL[e]		(–133, –175, –89)	{Gal}
HA-LSTc[e]		(–57, –154, 62)	{Gal}
LOL[f] site B		(29, 127, 39)	{Gal}
site C		(179, 121, –96)	{Gal}
PTO[g] site A		(–39, –162, 62)	{Gal}
site B		(–33, –104, –100)	{Gal}
site C		(32, –127, –55)	{Gal}

[a]ϕ_1 = C1–C2–O3′–C3′, ψ_1 = C2–O3′–C3′–H3′.
ϕ_2 = C1–C2–O6′–C6′, ψ_2 = C2–O6′–C6′–C5′, ω_2 = O6′–C6′–C5′–O5′.
[b]Wheat germ agglutinin complexed with a disialoglycopeptide (Wright and Jaeger, 1993).
[c]Wheat germ agglutinin complexed with 3′sialyl lactose (Wright, 1990).
[d]Monomer 5 (Merritt *et al.*, 1994).
[e]Influenza hemagglutinin (HA) complexes (Eisen *et al.*, 1997). 23 SL: 3′sialyl lactose, 26 SL: 6′sialyl lactose, LSTc: lacto-sero tetrasaccharide c.
[f]*Lathyrus ochrus* isolectin II (Bourne *et al.*, 1994).
[g]Pertussis toxin (Stein *et al.*, 1994).

points into the groove, and the glycerol side chain sticks out of the groove towards the solvent. In all other complexes, the sialic acid lies flat on the protein surface, and one side of the ring is not contacted by protein. The O4 atom is exposed, and the glycerol side chain points towards the protein and interacts with it. The NeuNAc-2 site of polyoma clearly belongs in the second class.

The conformations adopted by the oligosaccharides vary. Some examples are given in Table I. The conformation of the (α2,3)–glycosidic bond in the polyoma VP1 complex is similar to those seen in the complexes of wheat germ agglutinin and influenza hemagglutinin, and the hydrogen bond between NeuNAc and Gal that stabilizes this conformation is present in all these structures. NMR studies and energy calculations have shown that this conformation is one of three preferred ones (Breg *et al.*, 1989; Poppe *et al.*, 1989). Another NMR analysis concludes that there is a different preferred conformation (ϕ = –160°, ψ = –13°) for the NeuNAc–(α2,3)–Gal linkage (Sabesan *et al.*, 1991), but cholera toxin is so far the only protein–carbohydrate complex where this conformation has been observed (Merritt *et al.*, 1994).

There does not appear to be a clear preferred conformation for the NeuNAc–(α2,6) linkage. The torsion angles seen in the VP1 complex (ϕ = –84°, ψ = 162°, ω = 64°) are reasonably close to the values favored by NMR studies (Poppe *et al.*, 1992), and that conformation is also similar to the ones seen in the hemagglutinin–LSTc (Eisen *et al.*, 1997) and the wheat germ agglutinin–glycopeptide (Wright and Jaeger, 1993) complexes (Table I).

The rearranged segment at the N-terminus

The only significant structural differences between the assembled VP1 pentamer in the virus (Stehle and Harrison, 1996) and the unassembled VP1 pentamer fragment are the absence of the invading C-terminal arms and a rearrangement of residues 32–45 at the N-terminus (Figure 4). In the virus, the N-terminal residues 17–45 form a ‘clamp’ (strand B, helix 3A), which helps to fix the invading β-strand J from another pentamer, as shown in Figure 4A and B. In the recombinant pentamers, residues 35–45 run alongside strand I, with residues 35–40 forming a short and somewhat irregular β-strand (labeled ‘A_{new}’), which is linked to strand I by four main-chain hydrogen bonds (Figure 4A). The N-terminal segment 32–45 assumes an identical structure in all five copies. Residues 39–41 also take part in a small crystal contact that is very similar in all five monomers. It is therefore possible that the conformation of the N-terminal segment is merely a crystal-packing effect. We have recently found crystals of VP1 and a fragment of VP2/3, in which the pentamer packing is somewhat modified (X.Chen, T.Stehle and S.C.Harrison, unpublished). The conformation of the N-terminal segment in those crystals is the same as the one shown here. Thus, we believe that the observed difference between the conformation in the virion and the one in the free pentamer is significant, and that it is related to the presence or absence of the invading arm. We note that in the assembled capsid, the clamp interacts extensively and almost exclusively with the invading arm, and it is therefore highly unlikely that the N-terminal segment could retain its ‘clamp’ conformation in a free pentamer.

We believe that the alternative conformation of the N-terminal segment in the free pentamer has a role in the mechanism of assembly. VP1 pentamers can form capsid-like structures *in vitro* (Salunke *et al.*, 1986, 1989), and correct assembly requires that each VP1 monomer insert its emerging C-terminal arm into the β-sheet of a VP1 monomer in another pentamer. As illustrated in Figure 4B, it is geometrically possible for an arm emerging from one VP1 monomer to fold back and augment the BIDG β-sheet of that same monomer. Even if only one of the five arms were folded back in this manner, the ‘self-invasion’ would essentially poison the assembly process by producing partly assembled capsids that could not properly accept further pentamers. In the free pentamer, we find that the N-terminal segment interacts with the

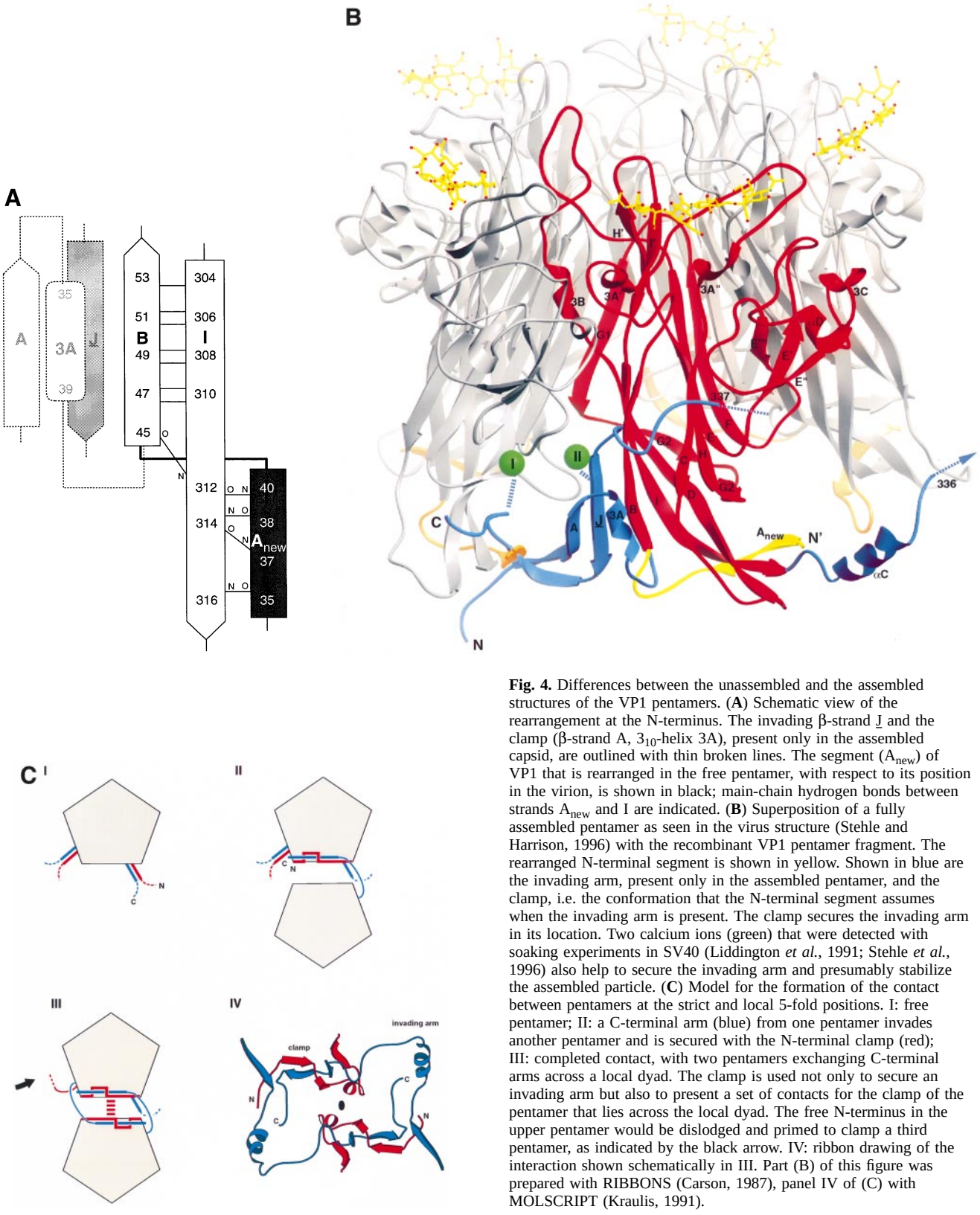

Fig. 4. Differences between the unassembled and the assembled structures of the VP1 pentamers. (**A**) Schematic view of the rearrangement at the N-terminus. The invading β-strand J and the clamp (β-strand A, 3_{10}-helix 3A), present only in the assembled capsid, are outlined with thin broken lines. The segment (A_{new}) of VP1 that is rearranged in the free pentamer, with respect to its position in the virion, is shown in black; main-chain hydrogen bonds between strands A_{new} and I are indicated. (**B**) Superposition of a fully assembled pentamer as seen in the virus structure (Stehle and Harrison, 1996) with the recombinant VP1 pentamer fragment. The rearranged N-terminal segment is shown in yellow. Shown in blue are the invading arm, present only in the assembled pentamer, and the clamp, i.e. the conformation that the N-terminal segment assumes when the invading arm is present. The clamp secures the invading arm in its location. Two calcium ions (green) that were detected with soaking experiments in SV40 (Liddington *et al.*, 1991; Stehle *et al.*, 1996) also help to secure the invading arm and presumably stabilize the assembled particle. (**C**) Model for the formation of the contact between pentamers at the strict and local 5-fold positions. I: free pentamer; II: a C-terminal arm (blue) from one pentamer invades another pentamer and is secured with the N-terminal clamp (red); III: completed contact, with two pentamers exchanging C-terminal arms across a local dyad. The clamp is used not only to secure an invading arm but also to present a set of contacts for the clamp of the pentamer that lies across the local dyad. The free N-terminus in the upper pentamer would be dislodged and primed to clamp a third pentamer, as indicated by the black arrow. IV: ribbon drawing of the interaction shown schematically in III. Part (B) of this figure was prepared with RIBBONS (Carson, 1987), panel IV of (C) with MOLSCRIPT (Kraulis, 1991).

C-terminal arm as it emerges from the subunit, and this interaction may help direct the arm away from the pentamer, thereby reducing the likelihood of self-invasion. Moreover, the N-terminal segment itself is prevented from forming a clamp to lock in an auto-invading C-terminal arm. Thus, we propose that the observed conformation of the N-terminus is part of a mechanism that helps ensure correct exchange of arms and increases the accuracy of assembly.

The assembly of polyomavirus is likely to be a 'self-controlled' process (Caspar, 1976) with the following properties (Stehle *et al.*, 1996). (i) There is a defined

Table II. Statistics for the complex and native data sets[a]

Resolution (Å)	R_{sym}[b] (%)	$<I>/<\sigma I>$	Refl. $>3\sigma$ (%)	Completeness (%)
25.0–4.4	3.6 (6.9)	22.5 (37.3)	99.0 (98.7)	81.4 (99.8)
4.4–3.5	4.6 (9.3)	23.1 (34.2)	98.3 (97.6)	91.0 (99.9)
3.5–3.1	5.0 (11.3)	23.6 (23.5)	95.1 (95.3)	96.9 (99.4)
3.1–2.8	5.8 (14.1)	19.4 (17.3)	92.4 (89.9)	97.6 (99.4)
2.8–2.6	6.7 (17.7)	15.4 (12.4)	87.9 (83.4)	96.9 (99.0)
2.6–2.4	7.6 (22.2)	12.3 (8.9)	82.8 (76.4)	97.1 (98.7)
2.4–2.3	8.8 (27.2)	10.0 (6.1)	76.9 (65.1)	96.3 (98.0)
2.3–2.2	10.8 (30.6)	7.7 (3.8)	70.8 (55.0)	94.6 (96.3)
2.2–2.1	11.9 (29.7)	5.5 (2.9)	62.0 (50.6)	90.2 (84.9)
2.1–2.0	15.2 (30.2)	3.8 (2.1)	55.6 (47.2)	82.7 (62.2)
2.0–1.9	23.2 (–)	2.1 (–)	48.3 (–)	56.5 (–)
Total	5.8 (10.6)	14.5 (14.9)	85.7 (72.1)	85.6 (83.2)

[a]Values for the native data set are shown in parentheses.
[b]$R_{sym} = \Sigma_h\Sigma_i|I_{hi}-I_h|/\Sigma_{hi} I_{hi}$ where h are unique reflection indices, I_{hi} are intensities of redundant, symmetry-related reflections of index h and I_h is the mean intensity for reflections of index h.

'assembly unit', in this case a VP1 pentamer. (ii) A specific nucleating structure initiates virion assembly, which then proceeds by addition of assembly units to a growing shell, rather than by coalescence of larger, stable intermediates. (iii) The bound conformation and bonding properties of a VP1 pentamer are determined uniquely and correctly by the characteristics of the site on the growing shell to which it adds. Recombinant VP1 pentamers form correct capsid-like structures without any other virion components (Salunke *et al.*, 1986, 1989). The information for accurate assembly is therefore contained in VP1 itself. We have proposed that the first step in assembly is the encounter of two pentamers, which interchange arms to form a contact resembling the one between a 5-coordinated and a 6-coordinated pentamer (Stehle *et al.*, 1996). As shown in Figure 4C, this contact includes interactions between two N-terminal clamp structures (shown in red) across a local dyad. Interchange of C-terminal arms (shown in blue) and rearrangement of N-terminal segments into clamps would be co-operatively stabilized by these additional clamp–clamp interactions, conferring particular stability on the contact in an initial assembly intermediate. Figure 4C also shows that arm invasion would influence the conformation of two N-terminal segments in the target pentamer—the one induced to clamp the invading arm and the one in the clockwise neighboring subunit. The latter would be dislodged from the pairing with its own C-terminus that we have just described in the present structure and primed to clamp an arm from a third pentamer (represented with a black arrow in Figure 4C), continuing the assembly process.

In summary, we believe that the conformation of residues 32–45 in the free pentamer is one of many structural details enhancing accuracy in virion assembly. As in protein folding, local 'native-like' free energy minima—such as self-invasion of a subunit by its C-terminal arm—are probably significant obstacles to achieving a correctly completed particle. The free pentamer is stable under physiological conditions *in vitro*, and it requires high salt concentrations to assemble into shells (Salunke *et al.*, 1989). Other sorts of assembly triggers, possibly including phosphorylation (Garcea *et al.*, 1985), are clearly relevant *in vivo*. Indeed, one candidate for phosphorylation, Thr113 in the CD-loop (T.Benjamin, personal communication), is immediately adjacent to residues 317–320 in the emerging C-terminal arm. Thus, the various contacts made by the N- and C-terminal arms, both in the virion and also as seen here in the pentameric assembly unit, may determine aspects of the assembly pathway as well as the stability of the final state.

Materials and methods

Protein purification and limited tryptic digestion

The VP1 fragment (residues 1–320) lacking the C-terminal arm was expressed and purified as described (Leavitt *et al.*, 1985; Garcea *et al.*, 1987). After attempts to crystallize this fragment had failed, we decided to carry out a limited proteolytic digestion. Incubation with trypsin (Sigma Corp.) for 6 h on ice (ratio 1:500) produced a slightly smaller, stable fragment that was not cleaved further. The tryptic digestion was halted after 6 h by adding BPTI, and the cleaved product was separated from the trypsin–BPTI complex by chromatography on Superdex-200 (Pharmacia Biotech). The VP1 pentamer fragment was collected and concentrated for crystallization. N-terminal sequencing of this fragment showed that trypsin had selectively cleaved after Lys31.

Crystallization

The N-terminally cleaved VP1 pentamers crystallized in space group $P3_121$ ($a = b = 221.5$ Å, $c = 99.8$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$) with one pentamer in the asymmetric unit. The solvent content is 65%. Crystals were grown using hanging drop vapor diffusion at room temperature. The drops contained 1.0 M ammonium phosphate pH 8.0, 2.5% ethanol and 8–10 mg/ml protein. The reservoir solution contained 2.0 M ammonium phosphate pH 8.0 and 5% ethanol. Crystals grew within 1–2 weeks up to a maximum size of $600\times600\times600$ μm^3. The presence of ethanol dramatically increased the size of the crystals and prevented the formation of 'crystal showers'. The crystals were collected into a harvesting solution that contained 2.2 M ammonium phosphate pH 8.0 and 5% ethanol. Prior to flash-cooling, the crystals were successively soaked for 30 min each in four depression wells which contained an increasing concentration of glucose (the cryoprotectant) in harvesting solution The glucose concentrations used were 5%, 10%, 20% and a final 33% (w/v). For the preparation of the oligosaccharide complex, a crystal was soaked for 1 h in harvesting solution that contained 33% glucose and 20 mM oligosaccharide (Oxford Glycosystems). The crystals were then scooped up with loops made of dental floss and flash-cooled by submerging them into liquid nitrogen.

Data collection and processing

Data were collected at the Cornell High Energy Synchrotron Source (F1 station) using a single flash-cooled crystal for each data set and a

Table III. Refinement results

	Complex	Native
Resolution (Å)	20.0–1.9	20.0–2.0
Number of reflections	188566	172107
R_{work} (%)	17.7	17.6
R_{free} (%)	20.5	19.9
r.m.s. bond deviation (Å)	0.007	0.008
r.m.s. angle deviation (°)	1.35	1.36
Number of protein atoms	11134	11134
average B factor (Å^2)	27.2	28.1
Number of water molecules	2127	2096
average B factor (Å^2)	39.9	39.1
Number of carbohydrate atoms	350	–
average B factor (Å^2)	51.5	–

wavelength of 0.91 Å. Data were recorded on Fuji imaging plates and scanned using a Fuji BAS2000 scanner (Fuji Inc., Japan). For the native data set, 114 high-resolution images with an oscillation range of 0.75° and 45 low-resolution images with an oscillation range of 2° were collected. For the complex data set, 76 high-resolution images were collected using a 1° oscillation range. Intensities were integrated and scaled with the HKL program package (Otwinowski, 1993) and converted into structure factor amplitudes with TRUNCATE (CCP4, 1994). The final data sets consist of 172 107 reflections (native data 20–2.0 Å) and 188 556 reflections (complex data set 20–1.9 Å). Data statistics are given in Table II.

Structure determination and refinement

The structure of the uncomplexed native VP1 pentamer was determined by molecular replacement as implemented in AMORE (Navaza, 1994). As search model, we used the polyoma VP1 pentamer coordinates which we had previously obtained through X-ray structure analysis of the complete virus particle at 3.65 Å resolution (Stehle *et al.*, 1994; Stehle and Harrison, 1996) The model was modified by deleting the N-terminal residues 17–31 and the C-terminal residues 321–383. The rotation function, calculated in several resolution ranges, clearly showed five maxima, corresponding to the five possible orientations of the pentamer. Translation searches were then carried out using one of the five equally strong rotation solutions and the possible space groups P321, $P3_121$, $P3_221$. Only the search conducted in $P3_121$ gave a clear solution, with an *R*-factor of 41.8% (20–6.0 Å). Rigid-body refinement in AMORE reduced the *R*-factor to 37.7%. Subsequent refinement was carried out in X-PLOR (Brünger *et al.*, 1987), with 2% of the data set apart prior to refinement to monitor the free *R*-factor (Brünger, 1992). A result of the 5-fold non-crystallographic symmetry is that the 'free' data set is no longer completely independent. Initial rigid body refinement, treating each VP1 monomer as a separate body, reduced the *R*-factor slightly, and we proceeded to carry out energy minimization and molecular dynamics simulation, at first including only data to 2.5 Å, then in later rounds extending the resolution to 2.0 Å. The model was rebuilt between rounds, using O (Jones *et al.*, 1991). Water molecules were located in ($2F_o$–F_c) maps with ARP (Lamzin and Wilson, 1993). Only water molecules with density $>1\sigma$ in the ($2F_o$–F_c) map were retained after each round. The refinement converged after eight rounds; the final model has excellent geometry and is in very good agreement with the experimental data (Table III). The location of the carbohydrate in the complex was determined by a difference Fourier map (Figure 2) using the phases from the refined native model to 2.0 Å. The oligosaccharide was built into that density, and water molecules that interfered with the ligand were deleted. The complex was then refined with X-PLOR, just as for the native model. The statistics are given in Table III. Both models contain residues 32–316 for all five monomers and residues 317–320 for monomer 2. Coordinates and structure factors have been deposited with the Brookhaven Protein Data Bank (accession numbers 1VPN and 1VPS); they are also available by e-mail (harrison@crystal.harvard.edu).

Acknowledgements

We acknowledge CHESS for access to the F1-beamline, and we thank the staff of CHESS and MacCHESS for help. We also thank X.Chen, T.Benjamin and D.L.D.Caspar for discussions. This work was supported by NIH-grant CA13202 (to S.C.H.) and by the Howard Hughes Medical Institute. S.C.H. is an investigator in the Howard Hughes Medical Institute.

References

Bauer,P.H., Bronson,R.T., Fung,S.C., Freund,R., Stehle,T., Harrison,S.C. and Benjamin,T.L. (1995) Genetic and structural analysis of a virulence determinant in polyoma VP1. *J. Virol.*, **69**, 7925–7931.

Bourne,Y., Mazurier,J., Legrand,D., Rouge,P., Montreuil,J., Spik,G. and Cambillau,C. (1994) Structures of a legume lectin complexed with the human lactoferrin N2 fragment, and with an isolated biantennary glycopeptide: role of the fucose moiety. *Structure*, **2**, 209–219.

Breg,J., Kroon-Batenburg,L.M.J., Strecker,G., Montreuil,J. and Vliegenthart,J.F.G. (1989) Conformational analysis of the sialyl α(2–3/6) N-acetyllactosamine structural element occurring in glycoproteins, by two-dimensional NOE 1H-NMR spectroscopy in combination with energy calculations by hard sphere exo-anomeric and molecular mechanics force field with hydrogen-bonding potential. *Eur. J. Biochem.*, **178**, 727–739.

Brünger,A.T. (1992) Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.

Brünger,A.T., Kuriyan,J. and Karplus,M. (1987) Crystallographic *R* factor refinement by molecular dynamics. *Science*, **235**, 458–460.

Cahan,L.D., Singh,R. and Paulson,J.C. (1983) Sialyloligosaccharide receptors of binding variants of polyomavirus. *Virology*, **130**, 281–289.

Carson,M. (1987) Ribbon models of macromolecules. *J. Mol. Graphics*, **5**, 103–106.

Caspar,D.L.D. (1976) Switching in the self-control of self-assembly. In *The Proceedings of the Third Annual John Innes Symposium*. pp. 85–99.

CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Connolly,M.S. (1983) Analytical molecular surface calculation. *J. Appl. Crystallogr.*, **16**, 548–558.

Eckhart,W. (1990) Polyomavirinae and their replication. In Fields,B.N. and Knipe,D.M. (eds), *Virology.* 2nd edn, Raven Press Ltd., New York, USA, pp. 1593–1607.

Eisen,M.B., Sabesan,S., Skehel,J.J. and Wiley,D.C. (1997) Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin-sialyloligosaccharide complexes determined by X-ray crystallography. *Virology*, **232**, 19–31.

Freund,R., Calderone,A., Dawe,C.J. and Benjamin,T.L. (1991) Polyomavirus tumor induction in mice: effects of polymorphisms of VP1 and large T antigen. *J. Virol.*, **65**, 335–341.

Fried,H., Cahan,L.D. and Paulson,J.C. (1981) Polyomavirus recognizes specific sialyloligosaccharide receptors on host cells. *Virology*, **109**, 188–192.

Garcea,R.L., Ballmer-Hofer,K. and Benjamin,T.L. (1985) Virion assembly defect of polyomavirus *hr-t* mutants: underphosphorylation of major capsid protein VP1 before viral encapsidation. *J. Virol.*, **54**, 311–316.

Garcea,R.L., Salunke,D.M. and Caspar,D.L.D. (1987) Site-directed mutation affecting polyomavirus capsid self-assembly *in vitro*. *Nature*, **329**, 86–87.

Jones,T.A., Zhou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Kawano,T., Koyama,S., Takematsu,H., Kozutsumi,Y., Kawasaki,H., Kawashima,S., Kawasaki,T. and Suzuki,A. (1995) Molecular cloning of cytidine monophospho-N-acetylneuraminic acid hydroxylase. Regulation of species- and tissue-specific expression of N-glycolyl neuraminic acid. *J. Biol. Chem.*, **270**, 16458–16463.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Lamzin,V.S. and Wilson,K.S. (1993) Automated refinement of protein models. *Acta Crystallogr.*, **D49**, 129–147.

Leavitt,A.D., Roberts,T.M. and Garcea,R.L. (1985) Polyomavirus major capsid protein, VP1. *J. Biol. Chem.*, **260**, 12803–12809.

Liddington,R.C., Yan,Y., Moulai,J., Sahli,R., Benjamin,T.L. and Harrison,S.C. (1991) Structure of simian virus 40 at 3.8 Å resolution. *Nature*, **354**, 278–284.

Luzzati,V. (1952) Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr.*, **5**, 802–810.

Merritt,E.A., Sarfaty,S., van den Akker,F., L'Hoir,C., Martial,J.A. and Hol,W.G.J. (1994) Crystal structure of cholera toxin B-pentamer bound to receptor G_{M1} pentasaccharide. *Protein Sci.*, **3**, 166–175.

Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.

Otwinowski,Z. (1993) Oscillation data reduction program. In Sawyer,L., Isaacs,N. and Burley,S. (eds), *Proceedings of the Daresbury Study Weekend: Data Collection and Processing*. Daresbury Laboratory, Warrington, UK, pp. 56–62.

Poppe,L., Dabrowski,J., von der Lieth,C.-W., Numata,M. and Ogawa,T. (1989) Solution conformation of sialosylcerebroside (G_{M4}) and its NeuNAc(α2,3) Galβ sugar component. *Eur. J. Biochem.*, **180**, 337–342.

Poppe,L., Stuike-Prill,R., Meyer,B. and Haalbeek,H.V. (1992) The solution conformation of sialyl-α(2,6)-lactose. Studies by modern NMR techniques and Monte Carlo simulations. *J. Biomol. NMR*, **2**, 109–136.

Ramachandran,G.N. and Sasisekharan,V. (1968) Conformation of polypeptides and proteins. *Adv. Protein Chem.*, **23**, 283–437.

Sabesan,S., Bock,K. and Paulson,J.C. (1991) Conformational analysis of sialyloligosaccharides. *Carbohydrate Res.*, **218**, 27–54.

Salunke,D.M., Caspar,D.L.D. and Garcea,R.L. (1986) Self-assembly of purified polyomavirus capsid protein VP1. *Cell*, **46**, 895–904.

Salunke,D.M., Caspar,D.L.D. and Garcea,R.L. (1989) Polymorphism in the assembly of polyomavirus capsid protein VP1. *Biophys. J.*, **56**, 887–900.

Sauter,N.S., Hanson,J.E., Glick,G.D., Brown,J.H., Crowther,R.L., Park,S.J., Skehel,J.J. and Wiley,D.C. (1992) Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography. *Biochemistry*, **31**, 9609–9621.

Stehle,T. and Harrison,S.C. (1996) Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. *Structure*, **4**, 183–194.

Stehle,T., Gamblin,S.J., Yan,Y. and Harrison,S.C. (1996) The structure of simian virus 40 refined at 3.1 Å resolution. *Structure*, **4**, 165–182.

Stehle,T., Yan,Y., Benjamin,T.L. and Harrison,S.C. (1994) Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. *Nature*, **369**, 160–163.

Stein,P.E., Boodhoo,A., Armstrong,G.D., Heerze,L.D., Cockle,S.A., Klein,M.H. and Read,R.J. (1994) Structure of a pertussis toxin–sugar complex as a model for receptor binding. *Nature Struct. Biol.*, **1**, 591–596.

Tooze,J. (1981) *DNA Tumor Viruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

Wright,C.S. (1990) 2.2 Å resolution structure analysis of two refined N-acetylneuraminyl-lactose-wheat germ agglutinin isolectin complexes. *J. Mol. Biol.*, **215**, 635–651.

Wright,C.S. and Jaeger,J. (1993) Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorin A. *J. Mol. Biol.*, **232**, 620–638.

Received on April 1, 1997; revised on May 20, 1997