Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment

Thilo Stehle*, Youwei Yan*, Thomas L. Benjamin† & Stephen C. Harrison‡

† Howard Hughes Medical Institute and
‡ Department of Biochemistry & Molecular Biology,
Harvard University, Cambridge, Massachusetts 02138, USA
§ Department of Pathology, Harvard Medical School, Boston,
Massachusetts 02115, USA

The polyomaviruses are non-enveloped, icosahedrally symmetrical particles with circular double-stranded DNA genomes. The outer shell of the virion contains 360 copies of viral protein VP1 (M, ~42K) arranged in pentamers. We report here the structure at 3.65 Å resolution of murine polyomavirus ("polyoma") complexed with an oligosaccharide receptor fragment. This structure has been determined using the previously described model of simian virus 40 (SV40). Although very similar in structure to SV40, polyoma has interesting biological differences. Cell-surface N-acetyl neuraminic acid (sialic acid) is required for polyoma infectivity, but not for SV40. Polyoma attaches to the surface of susceptible cells by stereospecific recognition of oligosaccharides terminating in (a2,3)-linked sialic acid. Studies of pathogenicity show that the specificity of viral binding to such oligosaccharides is an important determinant of the virus' ability to establish a disseminated infection and to induce tumours in the natural host. The complex described here shows how polyoma recognizes the receptor fragment and how strains with different receptor specificities can distinguish between alternative ligands. The results also suggest an explanation for the large disparity in pathogenicity exhibited by strains differing in only one amino-acid residue of VP1.

The receptor specificity of polyoma strains has been analysed using attachment to red blood cells (haemagglutination) as an assay. An (a2,3) linkage of sialic acid (NeuNAc) to galactose (Gal) is essential. Small plaque strains can tolerate attachment of a second sialic acid by (a2,6) linkage to N-acetyl galactosamime (GalNAc) in the following structure: NeuNAc-(a2,3)-Gal-(b1,3)-(NeuNAc-(a2,6)-)GalNAc. 'Large plaque' strains cannot bind to such a branched oligosaccharide. The critical difference between small and large plaque strains is the presence of glycine or glutamic acid, respectively, as residue 91 in VP1. Otherwise isogenic strains differ not only in their plaque morphology but also in their ability to cause tumours in mice. The small plaque strains produce few, if any, tumours; the large plaque strains are highly tumorigenic. Our crystals of the small plaque strain P16 diffract to 3.65 Å resolution, and we have been able to diffract 3'-sialyl lactose (NeuNAc-(a2,3)-Gal-β1,4-Glc) (Glc = glucose) into these crystals to obtain a complex. We note that, as expected, 6'-sialyl lactose does not bind. 3'-Sialyl lactose is a good model for the terminal parts of likely mammalian cell-surface ligands, such as NeuNAc-(a2,3)-Gal-β1,3-GalNAc, NeuNAc-(a2,3)-Gal-β1,3-GlcNAc or NeuNAc-(a2,3)-Gal-β1,4-GlcNAc. The structure determination is summarized in the caption to Fig. 1, which shows electron density for the trisaccharide. Details of the structure determination will be reported elsewhere. The refinement used data that extend to only moderate resolution. Nevertheless, the results obtained show that the model is in very good agreement both with experimental data (low "free" R-factor) and with standard geometry. The final averaged electron density map is of high quality, and the detailed similarity of the six independent VP1 coordinate sets in one icosahedral asymmetric unit shows that our model is sufficiently accurate for the interpretations that follow.

An overview of a polyoma subunit, with bound trisaccharide, appears in Fig. 2a. The view is essentially the same as the view of the SV40 subunit in ref. 4. The only significant conformational differences between VP1 of SV40 and polyoma are confined to surface loops connecting the β-strands of their 'jelly-roll' framework. Polyoma VP1 contains 22 more amino acids than SV40, of which 18 are found in the BC, DE, EF and HI loops (see ref. 4 and Fig. 2a for notation). The larger BC loop in polyoma can clearly be divided into two segments, labelled BC1 and BC2. The sialic acid lies in a shallow groove between loops BC1 and HI (Fig. 2b); the groove is closed off at one end by the tip of loop BC2 from the clockwise neighbouring subunit in the pentamer. The plane of the sugar ring lies roughly parallel to the walls of the groove. The carboxylate faces toward BC1, and the glycerol moiety faces toward DE. The galactose lies at the
FIG. 1 5-fold averaged electron density map for the 3-sialyl lactose.
The density is detailed enough to position and orient all three sugar rings. The map was calculated using (2Fobs - Fcalc) Fourier coefficients
between 1.2 and 3.65 Å resolution. Phases were calculated using the refined model and then further improved by 5-fold non-crystallographic
icosahedral averaging. The map is contoured at 1.0σ. For reasons of clarity, only contours within 1.5 Å of the model are plotted. The refined
model of 3-sialyl lactose is shown. The conformation of the trisaccharide is similar to the conformations seen in other protein/3-sialyl lactose
complexes. Map and model were obtained as follows: virus (Strain P16) was purified using a slightly modified version of the protocol described
in ref. 15. Crystals were grown by hanging drop vapour diffusion. Drops contained 6-8 mg ml⁻¹ virus, 10 mM HEPES pH 7.5, 0.25-
0.3 M sodium sulphate and 2.5% (v/v) glycerol; the reservoir contained 10 mM HEPES pH 7.5, 0.5-0.6 M sodium sulphate and 5% (v/v) glycerol.
Virus crystals grew within 2 to 4 weeks to a maximum size of 400 × 600 × 600 μm³. The crystal space group is I23 with a = 570 Å; there are 30 VP1 molecules in the asymmetric unit. Crystals were soaked for 20 h in collection buffer (0.65 M sodium sulphate, 5% (v/v) glycerol, 50 mM HEPES pH 7.5) containing 20 mM 3-sialyl lactose (Accurate Chemical & Scientific Corp., USA). Soaking did not noticeably perturb the crystal packing. Diffraction data were recorded on imaging plates (Fuji Inc., Japan) using an oscillation camera at the F1 synchrotron beamline of CHESS (wavelength 0.91 Å) and scanned using a Fuji

base of loop BC2, which forms an extended 'plateau' supported by part of loop EF. Loop DE of the anticlockwise neighbouring subunit approaches this plateau from the rear. The glucose ring lies above the plateau, but there are few if any van der Waals contacts to protein atoms. The thermal parameters for the glucose are higher than those for the sialic acid or galactose, consistent with the lack of strong constraint. Its orientation seems to be determined by a hydrogen bond between Glc-O3 and Gal-O5.

The shape of the binding groove is complementary to a preferred conformation of the NeuNac(a2,3)-Gal-linkage. This conformation is determined not only by potential steric hindrance for many choices of torsion angles about the glycosidic bond but also by a hydrogen bond from Gal-O2 to NeuNac-O6. A series of polar interactions between amino-acid side chains and groups of the sialic acid and galactose appears to orient the ligand further. Included in the description that follows are those hydrogen bonds for which the donor-acceptor distances are below 3.5 Å in all six independently refined coordinate sets. One key interaction appears to be a salt bridge between the carboxylate of the sialic acid and the guanidinium group of Arg 77. There are no negatively charged residues in the vicinity to neutralize the arginine, and it is possible that the electrostatic field generated by its uncompensated charge has a role in attracting and orienting the receptor. The nearby side chain of His 298 may enhance the charge field. The 'rear wall' of the groove is distinctly hydrophobic, with the side chain of Val 296 in direct contact with the plane of the sugar ring. Additional hydrogen bonds link the phenolic oxygen of Tyr 72 to the sialic acid amide group and His 298 to NeuNac-O4. There appears to be only one hydrogen bond to the galactose, between Asn 93 and Gal-O6. Gly 78 is probably also important for accommodating the galactose, because even a small side chain at this position would collide with Gal-O4.

Influenza virus also uses sialglycoconjugates as receptors, and the structures of the influenza haemagglutinin in complex with various α-sialosides have been determined. The primary binding site is a shallow pocket on the outer surface of the haemagglutinin. Only the sialic acid makes clear contacts, and both 3- and 6- sialyl lactose can bind. There is no homology with the interactions seen in the polycloma/sialyl lactose complex, but the number of hydrogen bonds is comparable. Both polycloma and influenza have quite low affinities for individual receptors (the dissociation constants are in the range of 1 mM) but when a virus attaches to a cell, more than one receptor is engaged, and a strong cooperative interaction occurs. By contrast, the tight binding of sugars by transporter-associated proteins involve very extensive hydrogen bond networks and very precisely complementary interaction.

It is possible to model related oligosaccharides into the polyoma site. In addition to the two pockets which accommodate the NeuNac(a2,3) Gal moey, we can discern a pocket for the
(α2,6) linked sialic acid of the branched-chain receptor (see Fig. 3). A preliminary analysis of crystals containing such a branched oligosaccharide shows that the second sialic acid is indeed located close to this pocket (T.S. and S.C.H., unpublished results). In one of the sterically favourable conformations of the branched linkage, the carboxylate of this sialic acid approaches the space where the side-chain carboxylate of glutamate 91 would be in large plaque strains. We therefore believe that coulombic interference can explain why large plaque strains do not bind branched-chain receptors.

Small plaque viruses, capable of recognizing branched as well as straight-chain sialyloligosaccharides, may bind more avidly.

FIG. 2 The receptor binding site of polyomavirus. a, Schematic drawing of a refined polyomavirus VP1 monomer together with its receptor analogue 3-sialyl lactose, shown as a ball-and-stick model. The loops that form the binding site are labelled. Residues 78 to 89 of the BC2 loop of the clockwise neighbour and residues 137 to 159 of the DE loop of the anticlockwise neighbour are also shown. The C-terminal arm that belongs to another pentamer and invades VP1 is shown in dark shading (see ref. 4). b, Enlarged view of the receptor binding pocket. The side chains of Tyr 72, Arg 77, Asn 93 and His 298 (in grey) as well as the trisaccharide (in yellow) are shown as ball-and-stick models. The Cα-atom of Gly 91 is marked with a black sphere. Hydrogen bonds are given as broken lines. a and b, Generated with MOLSCRIPT27. In the ball-and-stick models, oxygen atoms are red and nitrogen atoms are blue.

FIG. 3 Surface representation of a complete VP1 pentamer, with the 3-sialyl lactoses shown as ball-and-stick models. Four pockets (numbers 1–4) are evident. Pockets 1 and 2 accommodate sialic acid and galactose, respectively. Pocket 3 could accommodate the (α2,6)-linked sialic acid of a branched-chain receptor (see text). Although the function of pocket 4 has not yet been investigated, it is conceivable that it accommodates sugar moieties of still more complex oligosaccharide receptors. The arrow shows the position of the linkage to other sugars in complex oligosaccharides or to protein in short O-linked glycans. The figure has been generated with GRASP (A. Nicholls & B. Honig, Columbia University).

FIG. 4 View onto the virion surface, centred at a pentacoordinated VP1 pentamer (shown in green) on a strict icosahedral 5-fold symmetry axis. The VP1 molecules are represented with Cα-atom tracings, the corresponding trisaccharides are shown in a solid atom representation (carbon, yellow; oxygen, red; nitrogen, blue). The figure was generated with O26.
to cultured mouse cells by virtue of interaction with both types of oligosaccharides on the cell surface. It is therefore somewhat paradoxical that the small plaque strains, with broader oligosaccharide binding capacity, are effectively inhibited from spreading in the intact host1. A possible explanation for the lower pathogenicity of small plaque viruses would be that sialoglycoconjugates rich in branched chains are present in tissues of at least some mouse strains, and that such structures act effectively as "pseudoreceptors" or inhibitors specifically of these viruses. The relative densities of branched and straight-chain sialoglycoconjugates present on surfaces to which the virus can bind may also play a role. The distribution of VPI on the virus surface in principle allows simultaneous binding of many receptors without significant membrane distortion (Fig. 4), but the receptors would need to cluster within a region of 200-300 Å in diameter. The large plaque viruses may thus escape from inhibition by virtue of "ignoring" surfaces with a high density of branched sialoglycoconjugates.

Received 14 February; accepted 31 March 1994.


ACKNOWLEDGEMENTS. We thank S. Gambli for discussions and CHESS for time on the F3 beamline and the staff of CHESS and MacCHESS for their help. This work has been supported by grants from the National Cancer Institute of the NIH to T.L.B. and S.C.H. and by an EMBO postdoctoral fellowship to T.S.