Characterization of the Capsid Protein Glycosylation of Adeno-Associated Virus Type 2 by High-Resolution Mass Spectrometry†

Sarah Murray,3‡ Carol L. Nilsson,2 Joan T. Hare,3 Mark R. Emmett,1,2 Andrei Korostelev,3§ Heather Ongley,1 Alan G. Marshall,1,2 and Michael S. Chapman1,3*

Department of Chemistry & Biochemistry, Florida State University, Tallahassee, Florida 32306-43901; National High Magnetic Field Laboratory, Florida State University, 1800 E. Paul Dirac Dr., Tallahassee, Florida 32310-40052; and Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306-43803

Received 16 November 2005/Accepted 3 March 2006

Adeno-associated virus type 2 (AAV-2) capsid proteins have eight sequence motifs that are potential sites for O- or N-linked glycosylation. Three are in prominent surface locations, close to the sites of cellular receptor attachment and to neutralizing epitopes on or near protrusions surrounding the three-fold axes, raising the possibility that AAV-2 might use glycosylation as a means of immune escape or for preventing reattachment on release of progeny virus. Peptide mapping and structural analysis by Fourier transform ion cyclotron resonance mass spectrometry demonstrates, however, no glycosylation of the capsid protein for virus prepared in cultured HeLa cells.

Adeno-associated viruses (AAVs) are small single-stranded DNA paroviruses (36). There may be more than the eight serotypes originally identified, with humans or other primates as their primary hosts (13, 16). Although ~80% of the population is seropositive due to natural exposure to AAV (11, 34), AAV has not been associated with disease (3). This is one of the reasons that it is being developed as a vector for in vivo gene therapy (5, 33).

Parvoviruses are small (250 Å) unenveloped viruses in which a single-stranded DNA (ssDNA) genome is surrounded by a T=1 icosahedrally symmetric capsid containing 60 copies of the capsid protein. Parvoviruses contain three or four variants of the capsid protein, differing in length at the N terminus (36). In AAV type 2 (AAV-2), viral protein 3 (VP3; 533 amino acids) constitutes 80% (by mass) of the capsid (60). Alternative mRNA splicing gives variants VP2 and VP1, which are extended at the N terminus by 65 and 202 residues, respectively. Although present in the crystals, the VP1 and VP2 unique regions were not seen in the atomic structure (58)—it appears that the 533 residues common to VP1, VP2, and VP3 occupy symmetry-equivalent positions in the capsid. The VP1 unique addition encodes nuclear localization signals (22, 56) and a phospholipase A2 domain (15) that is likely needed for the virus to escape into the cytoplasm from the endosome (12a). The location of the VP1 unique region has been debated, but it looks as if can move from the inner to outer surfaces (12, 26, 59, 60).

The primary cellular receptor for AAV serotypes 1 to 3 is heparan sulfate (HS) proteoglycan (50). Coreceptors are likely also involved, possibly fibroblast growth factor receptor and integrin α5β1 (39, 40, 42, 49). AAV enters cells through endosomes (1) and is transported quickly into the nucleus (1, 44). Newly synthesized capsid proteins must be transported to the nucleolus for assembly of the DNA-containing virions (41), and specific binding of a nucleolus-targeting protein to VP2, and of nucleolin to AAV-2 virions has been reported (22, 41). Thus, the natural life cycle requires interactions with a number of other macromolecules in which glycosylation of the capsid proteins, if present, might be important.

Indeed, glycoproteins figure prominently in a number of other viruses, especially enveloped viruses, where they are often exposed on the outer surface. Mutations at the glycosylation sites often interfere with viral entry, infectivity, tissue specificity, or host range, implying roles in cell recognition, membrane fusion, and cell entry (8, 27, 43, 46, 53). Among nonenveloped viruses, glycosylation of structural proteins is less common. However, the fiber proteins of adenovirus, specifically types 2 and 5, are O glycosylated (6). The presence of the carbohydrate modulates the antigenicity in adenovirus (6). In rotavirus, it has been possible to select monoclonal antibody neutralization escape mutants (29) that have new sites of glycosylation in the epitope. Indeed, it has been suggested for human immunodeficiency virus type 1 (57) that glycosylation sites provide variability that allows the virus to escape immune detection of nearby conserved amino acids at the cellular-receptor binding site.

Our characterization of potential AAV-2 glycosylation sites started with a search through the sequence for likely motifs. It is possible to search for putative sites for N-linked glycosylation, for which there are several motifs (2), but for O-linked glycosylation, searching for motifs is more challenging (18, 24). The sequences of the capsid proteins were screened by use of

---

* Corresponding author. Mailing address: Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306-4380. Phone: (850) 644-8354. Fax: (850) 644-7244. E-mail: chapman@sb.fsu.edu.
† Supplemental material for this article may be found at http://jvi.asm.org.
‡ Present address: Medical Scientist Training Program, Albert Einstein College of Medicine, 1300 Morris Park Ave., Belfer 201, Bronx, NY 10461.
§ Present address: Molecular, Cell & Developmental Biology, 225 Sinsheimer, University of California, Santa Cruz, CA 95064.
Heparan sulfates are usually bound by basic residues (35). Electrostatic calculations from the atomic structure of AAV-2 (58) show that all are in prominent surface locations (Fig. 1). Of particular interest was the location that residues RGNR585–588 were at the center of a positively charged surface patch that was predicted to be the core of the receptor binding site (58), a prediction that was subsequently confirmed through mutagenesis (25, 38).

Examination of the location of the candidate sites on the atomic structure of AAV-2 (58) shows that all are in prominent surface locations (Fig. 1). Of particular interest was the location of NKS705–707 at the tip of the protrusions that surround each three-fold axis. It is in a location where glycosylation could impact both cell receptor binding and antigenicity. Heparan sulfates are usually bound by basic residues (35). Electrostatic calculations from the atomic structure showed that residues RGNR585–588 were at the center of a positively charged surface patch that was predicted to be the core of the receptor binding site (58), a prediction that was subsequently confirmed through mutagenesis (25, 38). Although far apart in the structure of an individual subunit, NNS496–498 and RGNR585–588 come together in the assembled capsid, as adjacent loops from neighboring three-fold related subunits (58). The candidate glycosylation sequence is at the tip of the protrusion whereas the receptor-binding site is on the inner side of the protrusion (58), but separated by only 3.1 Å, i.e., in immediate van der Waals contact and well within a likely footprint of the heparan sulfate receptor. Furthermore, NNS496–498 is within the major linear epitope of neutralizing antibody C37B, a monoclonal antibody that inhibits receptor attachment (55).

The remaining two candidate glycosylation sites (NGS382–384, and NKS705–707) are well separated within a subunit but come together on surface when contributed by adjacent subunits. They are on the opposite side of the RGNR585–588 receptor site, but only 21 to 25 Å away, still conceivably within a possible receptor footprint. Neither of these candidate glycosylation sites is within a neutralizing epitope yet mapped, but they are close. NGS382–384, is immediately adjacent in primary structure to epitope 2 of monoclonal antibody A20, and His134, which is a surface neighbor of both NGS382–384 and NKS705–707, is at the start of A20 epitope 1 (55, 58). With these candidate glycosylation sites having potentially significant impact upon both antigenicity and receptor binding, it became imperative to characterize glycosylation experimentally.

For experimental characterization, AAV-2 was prepared according to Xie et al. (60). The methods involve propagation of AAV-2 in cultured human (HeLa) cells, so it is not unreasonable to expect that glycosylation patterns should resemble those in a human host. There might be some differences based on cell type and with the higher in vitro propagation levels that could saturate posttranslational modification pathways. To reduce the potential for such heterogeneity, the virus was harvested 48 h postinfection (cf. 72 h). It is reasonable to expect that glycosylation patterns should resemble those in a human host. There might be some differences based on cell type and with the higher in vitro propagation levels that could saturate posttranslational modification pathways. To reduce the potential for such heterogeneity, the virus was harvested 48 h postinfection (cf. 72 h). It is reasonable to expect that at least some of the virus produced would have glycosylation mimicking that occurring naturally. Virus was purified by cesium chloride density gradient ultracentrifugation repeated three times.

Various enzymatic, chemical, and immunochemical analyses can reveal the presence of glycans (21, 28, 51). For preliminary screening, methods were chosen that had broad specificity—chemical oxidations detected with either fluorescence or immunoassay. For both assays, capsid proteins from denatured AAV-2 particles were electrophoretically separated on a 4 to 12% Bis-Tris gel (NuPage, Inc.). The Glycoprofile III fluorescent glycoprotein detection kit (Sigma, Inc.) was applied to a fixed gel according to the manufacturer’s instructions. Briefly,
Potential glycoconjugates were oxidized with a periodic acid solution, stained with dansyl hydrazide, destained, and imaged using a UV transilluminator with a yellow filter. EZBlue (Sigma, Inc) was then used to stain all proteins. The band for VP3 fluoresced with an intensity similar to that of ovalbumin. The weak fluorescent staining of VP1 and VP2 is not inconsistent with their ~10-fold lower concentration and glycosylation of a region common to VP1, VP2, and VP3.

The second assay used the digoxigenin (DIG) glycan detection kit (Roche, Inc) according to the manufacturer’s instructions. Potential carbohydrates were oxidized on the intact capsid with periodate caproic acid hydrazide and then labeled with digoxigenin. The particles were then denatured and separated electrophoretically. The proteins were transferred to a nitrocellulose membrane, then hybridized with an antidigoxigenin antibody conjugated to alkaline phosphatase so that activity could be measured colorimetrically. VP3 stained positively, but so did several unglycosylated proteins tested as negative controls. Thus, both chemical assays were giving inconclusive results. These types of tests are known to be difficult, with the real possibility of false positives. A more definitive method was sought, and the positive, if equivocal, chemical detection results motivated a mass spectrometric analysis.

Our choice would be liquid chromatography (LC)-Fourier transform ion cyclotron mass spectrometry (FT-ICR MS), taking advantage of the resources of the National High Magnetic Field Laboratory (NHMFL). There are important differences relative to the more familiar matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS). The preseparation of the samples by LC circumvents the problem of selective desorption/ionization, while the high resolution and mass accuracy provided by ICR permits the analysis of large glycopeptides (48). Also, the analysis of intact glycopeptides permits the assignment of specific sugar structures to a specific amino acid residue (19). Thus, it is preferable with FT-ICR MS to characterize directly the glycopeptides, rather than cleave off the carbohydrate as might be needed with MALDI-TOF MS, a particular challenge with O-linked glycosylation due to the lack of suitable enzymes.

One disadvantage of FT-ICR MS is the relatively large amount of sample needed. For the liquid chromatography mass spectrometry–mass spectrometry (LC-MS-MS) methods used to confirm the identity of potentially glycosylated peptides, larger quantities (up to 200 μg) would be needed than for MALDI-TOF MS. Such quantities are orders of magnitude greater than those normally prepared in culture for virological research. Replication of AAV-2 in HeLa cells is high yield relative to other cell types (37), and we took advantage of extensive past efforts to maximize production for crystallographic structure determination (60)—no further development of AAV-2 production in HeLa cells was required for the FT-ICR MS characterization. Samples for mass spectrometry were extracted from denatured gels and proteolytically fragmented. Stained VP3 and VP1 bands from eight lanes loaded with 8 μg denatured virus were excised from the gel and destained in 50% acetonitrile–25 mM ammonium bicarbonate until no color remained. Gel pieces were vacuum dried and then incubated overnight at 37°C with 15 μL of 10 μg/μL with trypsin, Asp-N, or Glu-C (Sigma-Aldrich, Inc.). Peptides were eluted from the gel pieces with 75% acetonitrile and 0.25% formic acid. VP1 required additional deionization with a ZipTip (Millipore, Inc). Samples were stored at −20°C. VP3, available in larger amounts, was analyzed first, followed by VP1.

Protein digests were dissolved in acetonitrile–0.5% formic acid (aqueous) at a concentration of approximately 10 μM and ionized by microelectrospray (9, 10). In order to obtain improved signal/noise ratio, some digests were purified with a miniaturized system (ZipTips, Millipore, Bedford, MA) according to the manufacturer’s instructions. Accurate mass analysis was performed with data obtained with a custom-built 7T or 9.4T Fourier transform ion cyclotron mass spectrometer (31, 45, 54). A one-dimensional nano-liquid chromatography (nano-LC) system (Eksigent, Livermore, CA) was interfaced with microelectrospray to the 7T FT-ICR MS and used for separation of protein digests over a 125-nl C18 cartridge (OptiPak, Optimize Technologies, Inc., Oregon City, OR). The nano-LC setup permitted efficient “peak parking” of proteolytic peptide ions. Peak-parked ions were stored waveform inverse Fourier transform (SWIFT) (17, 32) isolated in the
ICR cell and fragmented by infrared multiphoton dissociation (IRMPD) (30).

In total, 402 out of 533 amino acid residues, or 75.4% of the VP3 sequence, were mapped as unmethylated or methionine-oxidized (Fig. 3; and see Table S2 in the supplemental material). Oxidation of methionine residues frequently occurs during electrophoresis. The N-terminal peptide was not detected, although it was searched for against the mass data considering the possibility of N-terminal acetylation and/or terminal methionine truncation. The N terminus was also blocked when the possibility of N-terminal acetylation and/or terminal methionine oxidation (Fig. 3; and see Table S4 in the supplemental material). Oxidation of methionine residues frequently occurs during electrophoresis. The N-terminal peptide was not detected, although it was searched for against the mass data considering the possibility of N-terminal acetylation and/or terminal methionine truncation. The N terminus was also blocked when

FIG. 3. Coverage of the VP1 (clear) and VP3 (shaded) sequences by peptide mapping. Residues mapped are marked in boldface. For VP1, 599/733 (82%) of the amino acids in the sequence were covered. For VP3, 402/533 (75%) of the amino acids in the sequence were covered by the VP3 analysis or 415/533 (78%) were covered in either the VP3 or VP1 analysis (underlined).

For VP1, 599 out of a total of 733 amino acid residues, or 75.4% of the VP3 sequence, were mapped as unmethylated or methionine oxidized (Fig. 3; and see Table S2 in the supplemental material). Oxidation of methionine residues frequently occurs during electrophoresis. The N-terminal peptide was not detected, although it was searched for against the mass data considering the possibility of N-terminal acetylation and/or terminal methionine truncation. The N terminus was also blocked when subjected to Edman sequencing and thus is probably posttranslational modification. Residues mapped are marked in boldface.

For VP1, 599/733 (82%) of the amino acids in the sequence were covered. For VP3, 402/533 (75%) of the amino acids in the sequence were covered by the VP3 analysis or 415/533 (78%) were covered in either the VP3 or VP1 analysis (underlined).

In no case did the MS/MS data confirm the presence of modifying sugars. For VP1, several O-linked glycopeptide masses could be matched in the LC-MS data set from the trypic digest of VP3 (see Table S6 in the supplemental material), and 10 possible matches were found in the data from the Asp-N digest of the same protein (see Table S8 in the supplemental material). In neither set was the same glycan composition matched to the same peptide. In order to confirm the structures of the glycopeptides, the digested protein was subjected to LC-MS/MS analysis. Low sample volume and concentration necessitated the use of high-sensitivity nano-LC-MS. The Eksigent peak-parking function allowed the flow over the LC column to be greatly reduced, and thus the specific peptide ions of interest could be manipulated in the 7T FT-ICR (SWIFT isolated and IRMPD) for up to 20 min per digest fragment. The increased analysis time produced high-resolution tandem mass spectrometry (MS/MS) is necessary because of the risk of false-positive identifications. IRMPD is an important tool because it dissociates glycosidic bonds efficiently but preserves peptide bonds, thereby greatly simplifying the determination of the modifying glycan structure (19).

No N-linked glycopeptide masses could be matched with the GlycoMod tool to any of the experimental mass spectra submitted from either VP3 or VP1. In the case of O glycosylation, more matches are expected because of the large number of Ser and Thr residues in most proteins and because of the large number of possible glycan structures at each site. Five possible glycopeptide matches were found in the LC-MS data set from the trypic digest of VP3 (see Table S8 in the supplemental material), and 10 possible matches were found in the data from the Asp-N digest of the same protein (see Table S8 in the supplemental material). In neither set was the same glycan composition matched to the same peptide. In order to confirm the structures of the glycopeptides, the digested protein was subjected to LC-MS/MS analysis. Low sample volume and concentration necessitated the use of high-sensitivity nano-LC-MS. The Eksigent peak-parking function allowed the flow over the LC column to be greatly reduced, and thus the specific peptide ions of interest could be manipulated in the 7T FT-ICR (SWIFT isolated and IRMPD) for up to 20 min per digest fragment. The increased analysis time produced high-resolution tandem mass spectrometry (MS/MS) is necessary because of the risk of false-positive identifications. IRMPD is an important tool because it dissociates glycosidic bonds efficiently but preserves peptide bonds, thereby greatly simplifying the determination of the modifying glycan structure (19).

Thus, in spite of the presence of appropriate sequence motifs at suitably exposed locations on the surface of the capsid, it appears that at least the vast majority of AAV-2 particles are not glycosylated. A conservative estimate of the detection limit of FT-ICR MS is ~750 FM for a 15-μl sample. Gels loaded with about 10 μg of virus (protein plus DNA) yielded bands of about 5 μg VP3 and 1 μg VP1 and MS samples that were ~85 and 12 pM, respectively. Although negative mass spectrometry results cannot exclude the possibility that the capsid is ever glycosylated, no glycosylation is detected under conditions where 1% glycosylation of VP3 and 10% of VP1 should be detectable. It is therefore possible to define 1% as the upper bound for glycosylation of intact AAV-2.

This work was funded by NIH R01 GM66875 (M.S.C.), NSF DMR-00-84173 (A.G.M.), Florida State University, and the National High Magnetic Field Laboratory in Tallahassee, Fla.
REFERENCES


52. Reference deleted.