Glycoprotein I of herpes simplex virus type 1 contains a unique polymorphic tandem-repeated mucin region

Peter Norberg,¹ Sigvard Olofsson,¹ Mads Agervig Tarp,² Henrik Clausen,² Tomas Bergström¹ and Jan-Åke Liljeqvist¹

¹Department of Virology, University of Göteborg, Guldhedsgatan 10B, S-413 46 Göteborg, Sweden
²Department of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark

Glycoprotein I (gI) of herpes simplex virus type 1 (HSV-1) contains a tandem repeat (TR) region including the amino acids serine and threonine, residues that can be utilized for O-glycosylation. The length of this TR region was determined for 82 clinical HSV-1 isolates and the results revealed a polymorphic distribution of two to eight repeated blocks with a majority harbouring between two and four repeats. Assessment of the O-glycosylation capacity of an acceptor peptide (STPSSTTSTPSTT), representing two of the gI blocks, showed that the peptide was a universal substrate for O-glycosylation not only for the two most commonly expressed N-acetyl-D-galactosamine (GalNAc)-T1 and -T2 transferases, but also for the GalNAc-T3, -T4 and -T11 transferases. Immunoblotting of virus-infected cells showed that gI was exclusively O-glycosylated with GalNAc monosaccharides (Tn antigen). A polymorphic mucin region has not been described previously for HSV-1 and is a unique finding, as repeated blocks within gI homologues are lacking in other alphaherpesviruses.

Herpes simplex virus type 1 (HSV-1) is a DNA virus belonging to the family Herpesviridae and is classified in the subfamily Alphaherpesvirinae due to its capacity to establish latency in sensory neurons. HSV-1 encodes 11 glycoproteins, which are used in the early interactions between the virus and the target cell, in cell-to-cell spread and in evasion strategies to avoid the immune system of the host (Haarr & Skulstad, 1994). Glycoprotein I (gI) protein forms a non-convalently linked heterodimeric complex with glycoprotein E (gE), which has been shown to function as an Fc receptor binding monomeric and polymeric non-immune as well as anti-HSV IgG (Chapman et al., 1999; Dubin et al., 1990; Hanke et al., 1990; Johnson & Feenstra, 1987; Johnson et al., 1988). Another important function of the gE–gI complex is cell-to-cell spread in epithelial (Dingwell & Johnson, 1998) and neuronal tissue (Dingwell et al., 1995).

The gI gene of HSV-1 (strain 17) has been shown to contain a tandem repeat (TR) region (McGeoch et al., 1985). After DNA sequence analysis of the laboratory strains F, 17 and KOS321, as well as selected clinical HSV-1 isolates retrieved from different clinical entities (oral lesions, genital lesions and isolates inducing encephalitis), Norberg et al. (2004) recently showed that the TR region is polymorphic, varying from two to six or eight blocks, each containing 21 nt. The TR region encodes the amino acids serine, threonine and proline in a pattern characteristic for a peptide motif that might be utilized for O-linked glycosylation (Jentoft, 1990; Van Klinken et al., 1995). Glycoproteins containing a large number of O-glycosylated residues localized in a TR region are classified as mucins. Mucin-type O-linked glycosylation is initiated via the addition of an innermost N-acetyl-D-galactosamine (GalNAc) unit to the serine and threonine residues of the O-glycosylation motif, and this reaction is carried out by one or more isoenzymes of up to 20 human GalNAc transferase isoforms (Hassan et al., 2000; Ten Hagen et al., 2003). Although some overlapping may occur, these enzymes differ in kinetic properties, peptide specificity and tissue/cell expression patterns (Clausen & Bennett, 1996). No specific consensus sequence directing GalNAc incorporation into mucin regions is known. Whilst N-glycosylation only has one reducing terminal linkage in all eukaryotes, O-glycosylation has at least seven different classes of reducing terminal sugar linkage to serine and threonine (Gooley & Williams, 1994).

In the present study, we defined the length of the TR region in the gI gene in 82 clinical HSV-1 isolates received consecutively at the Department of Virology, University of Göteborg, Sweden. Twenty-five isolates were derived from oral lesions and 57 isolates from genital lesions. This distribution reflects the recent finding that HSV-1 is a common cause of first episodes of genital HSV in the western part of Sweden (Löwhagen et al., 2000). GMK-AH1 cell
cultures were infected with each isolate and grown until 100 % cytopathic effect was seen. Cell lysates were then used for the production of viral stocks and kept at −70 °C. Each viral stock representing a clinical HSV-1 isolate was passaged twice in cell culture before examination. In a recent study, the TR region was shown to be genetically stable after three passages in GMK-AH1 cells (Norberg et al., 2004).

The TR region of the gI gene was analysed using a length-specific PCR technique. HSV-1 DNA was prepared using a QIAmp blood kit (Qiagen). The forward primer 5′-CGGGCTATGGCTCTGGACCC-3′ (nt 530–550, strain 17) and reverse primer 5′-ATGGGGGTTCGTGGTGA-3′ (nt 766–746, strain 17) were used and samples were amplified as described previously (Norberg et al., 2004). PCR products were separated by Metaphor agarose gel electrophoresis using PCR products from clinical isolates containing two to six or eight repeated blocks as size markers. In an initial evaluation, we verified that the length of the PCR products corresponded to the number of TR regions in the clinical isolates by DNA sequencing using an ABI Prism 310 genetic analyser (Applied Biosystems). The distribution of the number of TR regions for 82 clinical HSV-1 isolates was as follows; 18 isolates (22 %) contained two repeats, 24 isolates (29.3 %) contained three repeats, 20 isolates (24.4 %) contained four repeats, 12 isolates (14.6 %) contained five repeats, six isolates (7.3 %) contained six repeats and two isolates (2.4 %) contained eight repeats. The majority of the clinical HSV-1 isolates (75.6 %) displayed two, three or four repeats. The laboratory HSV-1 strains KOS321, 17 and F displayed two, three and five repeated blocks, respectively. No association was found between the number of repeated blocks and localization of the clinical lesions (oral or genital).

First, we determined the ability of the gI TR region to function as a substrate for glycosylation using a selection of recombinantly expressed human GalNAc transferase isoforms (GalNAc-TN, where N defines the numerical designation of a given enzyme). In vitro assessment of the O-glycosylation capacity of the TR region using polypeptide GalNAc transferase assays was carried out essentially as described previously using soluble, secreted, purified recombinant human GalNAc-T1, -T2, -T3, -T4 and -T11, expressed in insect cells (Bennett et al., 1998; Wandall et al., 1997). The TR region in gI encodes three similar but not identical blocks of residues (STPSTTT, STPSTTI or PAPSTTI). The first block is the most abundant (Norberg et al., 2004) and was here used in duplicate (shown in italic) as an acceptor substrate within the peptide KPNPASTPSSTTSSTPTTTPAPK. The peptide was synthesized by Chiron Mimotopes and purified by the manufaturer to 99 % purity. Briefly, assays were performed in 25 µl total reaction mixture containing 25 mM sodium cacodylate (pH7.4), 10 mM MnCl2, 0.25 % Triton X-100, 200 µM of the monosaccharide UDP-14C-GalNAc (3700 c.p.m. nmol−1; Amer sham) for kinetic analysis (Table 1) or 2 mM UDP-GalNAc for mass spectrometry analysis (Fig. 1), 50 or 200 µM acceptor peptide for assays with GalNAc-T1, -T2, -T3, -T4 and -T11, and 0.25 mU of each purified GalNAc transferase. Soluble, secreted forms of the human GalNAc transferases analysed were expressed in High Five cells, grown in serum-free media and purified to near homogeneity with specific activities of 0.5–2.35 U mg−1 measured using peptides derived from the TR regions in the human mucin genes MUC1, MUC2 and MUC7, as described previously (Wandall et al., 1997). Enzymic activity was measured by incubating 0.25 mU enzyme for 10 min at 37 °C followed by Dowex-1 chromatography and liquid scintillation counting.

The molecular mass of each peptide was analysed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. Terminal results of the spectra obtained after a 24 h incubation for all enzymes are presented in Fig. 1, and a kinetic analysis of the initial reaction velocities for GalNAc-T1 to -T4 and T11 are shown in Table 1. The results showed that the TR region in gI is a substrate for O-glycosylation. GalNAc-T1 and -T2 were able to add six to seven GalNAc residues to the gI peptide as shown by the major peaks in the mass spectrometry analysis (Fig. 1). GalNAc-T3 was consistently found to cause the addition of seven GalNAc residues to the peptide. GalNAc-T4 varied from the other enzymes in as much as the initial rate was considerably lower. On the other hand, long-term incubation with the enzyme resulted in a higher number of GalNAc residues being added (between nine and 11 residues) compared with GalNAc-T1 to -T3. Finally, GalNAc-T11 added four to five residues (Table 1 and Fig. 1). Taken together, these results suggest that the repeated sequence in gI serves as a substrate for O-glycosylation and that multiple GalNAc transferase isoforms can participate in this process.

The lengths of the gI protein for clinical HSV-1 isolates containing two and eight repeats were analysed. Antigen for immunoblotting was prepared by infecting monolayers

### Table 1. Summary of MALDI-TOF mass spectrometry evaluation of GalNAc-TN activity expressed as the number of GalNAc residues incorporated into the gI peptide and the enzymic activity at two different concentrations of the peptide

<table>
<thead>
<tr>
<th>GalNAc</th>
<th>No. glycosylated sites*</th>
<th>Enzymic activity† (pmol min⁻¹) at a peptide concentration of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>50 µM</td>
</tr>
<tr>
<td>T1</td>
<td>6–7</td>
<td>33</td>
</tr>
<tr>
<td>T2</td>
<td>6–7</td>
<td>11</td>
</tr>
<tr>
<td>T3</td>
<td>7</td>
<td>68</td>
</tr>
<tr>
<td>T4</td>
<td>9–11</td>
<td>5</td>
</tr>
<tr>
<td>T11</td>
<td>4–5</td>
<td>17</td>
</tr>
</tbody>
</table>

*Observed range of incorporated GalNAc residues. The results from mass spectrometry analysis are indicated as major peaks (Fig. 1).
†Enzyme (20 µU, 0.2–0.4 µg) was incubated for 10 min at 37 °C. Enzymes were standardized using a MUC7 21mer peptide.
of HEL-R66 cells with an m.o.i. of 10. When 100% cytopathic effect was seen, the cells were harvested and lysed in Tris-buffered saline and 1% NP-40, followed by ultrasonication. The samples were mixed with sample buffer containing 2% SDS, 5% mercaptoethanol and subjected to PAGE using a 10% Bis/Tris gel (Novex) and 3-\(N\)-morpholino)propane sulfonic acid as running buffer. Proteins were electrotransferred to an Immobilon-P transfer membrane (Millipore). A monospecific rabbit anti-gI serum (UP 1928), kindly provided by Harvey Friedman (PA, USA), was diluted 1:200 and incubated overnight. Peroxidase-labelled swine anti-rabbit IgG (Dako) at a 1:100 dilution was used as conjugate with 4-chloro-1-naphthol as substrate. The gI protein containing two repeated blocks has a predicted molecular mass of 40.7 kDa, whilst gI containing eight blocks has a predicted molecular mass of 44.8 kDa (14 additional amino acids per two repeats). Fig. 2(a) shows the different lengths of the gI proteins, which showed a difference in molecular mass from the predicted sizes of approximately 15 kDa. The gI gene was sequenced for the isolates used and revealed that the three potential \(N\)-glycan sites as well as serine and threonine residues localized in the flanking regions were conserved. Instead, due to the high frequency of the amino acids serine, threonine and proline in the TR region, the molecular shift is most likely explained by the addition of O-glycans attached to the repeated amino acid motif.

We further explored whether gI derived from HSV-1-infected HEL-R66 cells was O-glycosylated. To define typical O-linked glycans on gI, we performed blot experiments using a mAb (5F4) specific for GalNAc-serine/threonine (Tn antigen) (Mandel et al., 1991). Undiluted hybridoma supernatant was incubated overnight at 4°C with 15μl virus-infected cell lysate. Alkaline phosphatase-labelled

![Fig. 1. MALDI-TOF mass spectra of an in vitro-glycosylated peptide, KPNPASTPSTTSTPSTTPPAPK, representing two repeats in gI (in italics). The figure shows the results after 24 h using the GalNAc-T1 to -T4 and T11 transferases. The different peaks represent the number of GalNAc residues incorporated.](http://vir.sgmjournals.org/1685)

![Fig. 2. (a) The gI protein from clinical HSV-1 isolates harbouring two and eight TR regions was separated from lysates of virus-infected HEL-R66 cells by SDS-PAGE and detected using rabbit anti-gI antibody (UP 1928). (b) Reactivity of mAb 5F4 specific for GalNAc monosaccharides attached to serine/threonine (Tn antigen). (c) PNA assay with affinity for disaccharide Gal(β1–3)GalNAc attached to serine/threonine (T antigen). Molecular mass markers are indicated in kDa.](http://vir.sgmjournals.org/1685)
anti-mouse antibody was used as conjugate and BCIP/NBT as substrate. The mAb identified a single band of gI for each of the isolates (Fig. 2b), where the electrophoretic mobility corresponded to the long and short forms of gI, respectively. The band representing the HSV-1 isolate containing eight tandem repeats stained more intensely than its shorter counterpart, indicating that the longer gI form contained more GalNAc monosaccharides, as expected.

Next, we used peanut agglutinin (PNA) with affinity to Gal(β1–3)GalNAc-serine/threonine (T antigen) (Goldstein & Hayes, 1978). Biotinylated PNA was incubated overnight at 4 °C with 15 μl virus-infected cell lysate. Streptavidin–alkaline phosphatase and BCIP/NBT were used for detection. The PNA assay visualized two more-slowly migrating bands, which were identical for both HSV-1 isolates and clearly separated from the gI proteins (Fig. 2c). In a separate experiment, we demonstrated that these two bands could be stained with a gC-1-specific antibody (C2H12; Sjoblom et al., 1992) (data not shown), which is in line with previous observations that gC-1 is heavily covered with PNA-binding O-linked glycans (Gudrunström et al., 1987). This result showed that the inability of gI to accept O-linked glycans larger than monosaccharides (Tn) was not due to the absence of relevant glycosyltransferases in the infected cells. Thus, gI is densely covered with GalNAc monosaccharides attached to individual serine/threonine residues. In addition, these O-linked monosaccharides were not further extended to the T (disaccharide) level, despite the fact that HEL-R66 cells are capable of performing such modifications. This finding is intriguing and two possible explanations can be considered: the clustering of GalNAc units in gI may cause steric hindrance for elongation of additional sugar moieties, or the gI protein may contain a molecular signal that precludes further transport from the cis-Golgi, the compartment where the initial glycosylation is initiated.

A prerequisite for O-linked glycosylation in a specific region is the presence of serine, threonine and proline residues at a relatively high frequency (Jentoft, 1990; Van Klinken et al., 1995). Although defined consensus acceptor sequence motifs for mucin-type O-linked glycosylation has not been identified, analysis of the in vitro O-glycosylation capacity of recombinant GalNAc transferase isoforms with short peptide sequences covering potential acceptor sites has been found to be highly predictive of in vivo-utilized sites (Defrees et al., 2006). The results of the present study thus indicate that the repeated amino acid sequences in gI are efficient signals for O-linked glycosylation and are a substrate not only for the two most commonly expressed specificity types (GalNAc-T1 and -T2), but also for GalNAc-T3, -T4 and -T11. HSV-1 infects the skin, cervix and neurons. Expression of the GalNAc transferases is variable among different cell types and during differentiation. An earlier study has shown that the distribution of GalNAc-T1 to -T3 differs among the layers in the human skin, cervix and the epithelium of mouth in that GalNAc-T2 was expressed mostly in the undifferentiated basal and parabasal cells and GalNAc-T1 in the upper layer in the non-keratinized layer. In contrast, GalNAc-T3 was expressed in all cell layers in equal amounts (Mandel et al., 1999). As HSV-1 replicates in the basal and parabasal cells, it is likely that GalNAc-T2 and -T3 are the most important transferases for O-glycosylation of gI during in vivo HSV-1 infection. The distribution of the transferases in the central nervous system has not been investigated thoroughly, but several GalNAc transferase isoforms including GalNAc-T4 have been shown to be expressed in low amounts (Bennett et al., 1998, 1999).

The HSV-1 genome contains several TR regions. Most of these sequences are non-coding and flank the unique long and short regions (McGeoch et al., 1985, 1986, 1988; Perry & McGeoch, 1988). TR regions in coding sequences are sparse and the polymorphic region described here encoding a mucin tract is a unique finding in the HSV-1 genome. To our knowledge, other alphaherpesviruses lack gl homologues containing a polymorphic mucin region, suggesting that HSV-1 acquired this region after its division from HSV-2. This region may therefore have a function exclusive to HSV-1. Mucins are a family of highly glycosylated proteins, usually secreted or present on apical cell membranes of human epithelial cells. A specific characteristic of human mucins is the polymorphism of O-glycosylated TR regions generating highly variable sizes of glycoproteins. Mucins present on the cell surface may act as a barrier between the cell surface and the surrounding environment, and are involved in protection and lubrication of epithelial surfaces against micro-organisms, toxins and proteolytic attack. Furthermore, the extracellular portion of the mucin protein can be involved in signal transduction via ligand binding, as described for the human mucin MUC1, which binds to intracellular adhesion molecule 1 (Singh & Hollingsworth, 2006).

The gl protein is a type I protein with a single transmembrane domain and a cytosolic tail at the C terminus. No function has been described for the gl protein alone. The gE–gl complex binds monomeric IgG, and aa 128–145 in gl (strain 17), located more N terminally than the TR region (aa 208–228, are required for the Fc receptor function (Basu et al., 1997). As O-glycans are proposed to affect the adoption of a rigid and extended structure by the protein (Olofsson, 1992), it is possible that gl–gE complex formation and binding of monomeric IgG is modulated by the variable length of the mucin region for different HSV-1 isolates. The variability of the mucin region described here for gl may therefore have consequences for pathogenesis, tropism and transmission of HSV-1 and has to be considered when functions of gl alone or in complex with gE are studied.

Acknowledgements

We thank Maria Johansson for excellent technical assistance and Harvey Friedman, PA, USA, who kindly provided rabbit anti-gl
serum. This work was supported by grants from the Medical Society of Göteborg, Swedish Medical Research Council (grant no. 11225), the LUA foundation at Sahlgren’s Hospital, the Swedish Society for Medical Research and the Danish Medical Research Council.

References


