Analysis of the Pre-S2 N- and O-Linked Glycans of the M Surface Protein from Human Hepatitis B Virus*

(Received for publication, December 14, 1998, and in revised form, February 12, 1999)

Sigrid Schmitt‡, Dieter Glebe§, Kim Alving§, Tanja K. Tolle§, Monica Linder, Hildegard Geyer, Dietmar Linder, Jasna Peter-Katalinic, Wolfram H. Gerlich§, and Rudolf Geyer¶

From the Institute of Biochemistry, University of Giessen, Friedrichstrasse 24, D-35392 Giessen, Germany, the ¶Institute of Medical Virology, University of Giessen,Frankfurter Str. 107, D-35392 Giessen, Germany, and the §Institute of Medical Physics and Biophysics, University of Münster, Robert-Koch-Straße 31, D-48149 Münster, Germany

The surface antigen of hepatitis B virus comprises a nested set of small (S), middle (M), and large (L) proteins, all of which are partially glycosylated in their S domains. The pre-S2 domain, present only in M and L proteins, is further N-glycosylated at Asn-4 exclusively in the M protein. Since the pre-S2 N-glycan appears to play a crucial role in the secretion of viral particles, the M protein may be considered as a potential target for antiviral therapy. For characterization of the pre-S2 glycosylation, pre-S2 (glyco)peptides were released from native, patient-derived hepatitis B virus subviral particles by tryptic digestion, separated from remaining particles, purified by reversed-phase high performance liquid chromatography, and identified by amino acid and N-terminal sequence analysis as well as matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). Pre-S2 N-glycans were characterized by anion exchange chromatography, methylation analysis, and on target sequential exoglycosidase digestions in combination with MALDI-TOF-MS, demonstrating the presence of partially sialylated diantennary complex-type oligosaccharides. In addition, the pre-S2 domain of M protein, but not that of L protein, was found to be partially O-glycosylated by a Gal(β1-3)GalNAc, Neu5Ac(O2–3)Gal(β1–3)GalNAc, or GalNAco-residue. The respective O-glycosylation site was assigned to Thr-37 by digestion with carboxypeptidases in combination with MALDI-TOF-MS and by quadrupole time-of-flight electrospray mass spectrometry. Analytical data further revealed that about 90% of M protein is N-terminally acetylated.

Hepatitis B virus (HBV),1 belonging to the virus family hepadnaviridae, is an important etiological agent of acute and chronic liver disease (1, 2). Chronic HBV infection may lead to liver cirrhosis and hepatocellular carcinoma, which result in about 1 million deaths per year worldwide. The virus replicates in the liver and is secreted in large amounts of up to 1010 particles/ml into the blood (3). In addition to 42-nm DNA containing virions, infected hepatocytes produce subviral, noninfectious 22-nm spherical or filamentous particles in vast excess. The envelopes of virions and subviral particles contain varying amounts of three related HBV-encoded (glyco)protein species termed large (L), middle (M), and small (S) proteins, which are together referred to as HBV surface antigen (HBsAg). S protein is the major component of virions and both spherical and filamentous HBsAg particles, while filaments and virions contain more M and, in particular, more L proteins than spheres (4, 5).

All envelope proteins are produced from a single open reading frame (see Fig. 1A) by the use of three different translation start sites, dividing this open reading frame into three domains: the amino-terminal pre-S1 domain, which occurs exclusively in the L protein; the pre-S2 domain, which is present in both M and L proteins and forms the amino-terminal end of the M protein; and the S domain, which is common to S, M, and L proteins. All proteins possess a potential N-glycosylation site at Asn-146 of the S domain, which, however, is only partially utilized. Hence, the proteins exist in two isomeric forms, being either glycosylated or nonglycosylated in this position, and migrate as doublets in SDS gel electrophoresis (Fig. 1B). The second potential N-glycosylation site, present at Asn-42 of the pre-S2 domain, is solely used in the M protein but not in the L protein (4). In addition to N-glycosylation, pre-S2 domains of HBV M protein expressed in mammalian cell culture (6–8) or of M protein of the related woodchuck hepatitis virus (WHV) (9) have been reported to be O-glycosylated in as yet unidentified positions.

HBV envelope (glyco)proteins are functionally important during the viral life cycle. L protein mediates binding to human hepatocytes of the virus via its pre-S1 domain (10, 11). Furthermore, L protein has been shown to be important for virion envelopment and secretion (12). In order to fulfill these functions, the pre-S domain of the L protein has a dual topology. During and directly after translation, it is located at the cytosolic side of the endoplasmic reticulum and, hence, within the virion after budding. Therefore, L is not N-glycosylated in its

---

* This work was supported by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 535, Projects Z1 (to R. G., and D. L.) and A2 (to W. H. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be identified "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† This paper is in partial fulfillment of the requirements for the degree of Dr. rer. physiol at Marburg University.

‡ To whom correspondence should be addressed: Institute of Biochemistry, University of Giessen, Friedrichstrasse 24, D-35392 Giessen, Germany. Tel.: 49-641-99-47400; Fax: 49-641-99-47409; E-mail: Rudolf.Geyer@biochemie.med.uni-giessen.de.

§ The abbreviations used are: HBV, human hepatitis B virus; BSA, bovine serum albumin; ESI, electrospray ionization; HBsAg, hepatitis B virus surface antigen; Hex, hexose; HexNAc, N-acetylhexosamine; HPAEC, high pH anion exchange chromatography; MALDI, matrix-assisted laser desorption/ionization; TOF, time-of-flight; MS, mass spectrometry; QTOF, quadrupole time-of-flight; HPLC, high-performance liquid chromatography; rHPLC, reversed-phase HPLC; S, M, and L proteins, small, middle and large hepatitis B surface proteins, respectively; sialosyl T-antigen, Neu5Acα2–3Gal(β1–3)GalNAc-Ser/Thr; T-antigen, Thomsen-Friedenreich antigen Galβ1–3GalNAc-Ser/Thr; Tα-antigen, GalNAcα1–2GalNAc-Ser/Thr; WHV, woodchuck hepatitis B virus; PAGE, polyacrylamide gel electrophoresis.

1 Amino acid residues derived from the pre-S2 region are underlined.
pre-S region. Later, L molecules change, in part, their topology by an unknown mechanism, resulting in a surface exposure of pre-S1 and pre-S2 at HBsAg particles and virions (13). In contrast, the pre-S2 domain of M protein is translocated cotranslationally to the endoplasmic reticulum lumen and is, thus, N-glycosylated. The role of M protein was not clear for a long time. Recent studies provided evidence, however, that M protein glycosylation at Asn-4 of the pre-S2 domain as well as subsequent trimming of this N-glycan play an important role in the secretion of HBV virions (8, 14–19). Prevention of M protein N-glycosylation by either disruption of the Asn-4-X-Ser-6 sequon (16) or by tunicamycin treatment of HBV-expressing cells (15) suppressed secretion of HBV particles. Likewise, treatment of cell cultures with inhibitors of the oligosaccharide-trimming enzymes α-glucosidase I and II similarly impaired virion secretion (17, 18). The pre-S2 N-glycan mediates an association of the M protein with the chaperone calnexin, whereas the N-glycan linked to Asn-146 of the S domain is not involved in calnexin binding (8, 16). It is concluded that proper folding and trafficking of the M glycoprotein, assisted by calnexin in a carbohydrate-dependent manner, may play a crucial role in the assembly of virions, whereas secretion of subviral particles is not prevented by glycosylation- and trimming-inhibitors (15–19). The pre-S2 N-glycan of the M protein may be considered as a promising target for antiviral therapy of hepatitis B, because viremia can be suppressed by α-glucosidase I inhibitors in experimentally infected woodchucks (18). Previous studies on the structure(s) of the N-glycan at Asn-4 of HBV M protein provided contradictory data proposing either partially fucosylated, complex-type (16, 20) or high mannose- and hybrid-type species (21). The presence of terminal mannose would have been biologically interesting, since a genetic linkage between polymorphism of the mannose-binding protein and the persistence of HBV infection was observed (22).

The pre-S1 and pre-S2 domains of HBsAg display numerous T- and B-cell epitopes (23, 24), which are capable of inducing neutralizing antibodies and immune protection (25, 26). Hence, L and M proteins are also attractive candidates for the development of improved vaccines against HBV, which might (a) override nonresponsiveness to the standard HBV vaccine consisting only of S protein (27), (b) allow immunotherapeutic treatment of chronic HBV infections (28), and (c) prevent selection of escape mutants with mutations in the S protein (29). Several candidate hepatitis B vaccines containing pre-S2 sequences have been developed that are produced in transfected cell cultures (27, 28, 30–33). The usefulness of these vaccines is not completely proven yet, particularly because the anti-pre-S response induced by these vaccines in human recipients is rather weak.3 One reason for this observation may reside in modifications of the pre-S domain. Since naturally generated HBsAg is the target of neutralizing antibodies in vivo, knowledge on post-translational modifications of the pre-S2 sequence is important for the design and evaluation of future hepatitis B vaccines. Current data on N- and O-glycosylation or phosphorylation (8, 34) of the pre-S2 region were obtained from trans-

3 W. H. Gerlich, unpublished observations.
FIG. 2. Separation of tryptic peptides derived from native HBsAg particles. Peptides were fractionated by rHPLC on a C<sub>18</sub>-column (5 μm; 4.6 × 250 mm), using a gradient of acetonitrile in 0.1% aqueous trifluoroacetic acid as indicated. The N-terminal pre-S2 glycopeptide from the M protein (pre-S2 positions 1–16) eluted within a time interval of about 5 min, marked by bracket Ia. The corresponding peptide of the L protein (−6 to 16; see Fig. 1C) was not glycylated and eluted mainly as peak 11 (bracket Ib). The unglycosylated C-termin al tryptic peptides of the pre-S2 domains (positions 19–48) from M and L proteins eluted as peak 19 (bracket IIa), whereas bracket IIb indicates the elution positions of O-glycosylated peptides 19–48. For further explanations, see also Table I.

Experimental Procedures

Isolation of HBsAg—Subvirial spherical and filamentous particles were purified from the sera of two chronically HBV-infected donors, genotype D (HBsAg subtype ayw2), with a virus titer of about 6 × 10<sup>10</sup>/ml and a HBsAg concentration of 100 μg/ml. The genotype and the HBsAg subtype were determined by sequencing of the viral genomes as described previously (35). 18 ml of serum were ultracentrifuged in a discontinuous sucrose density gradient (15, 25, 35, 45, and 60% (w/w)) layered for further purification between a KBr density gradient ranging from 1.16 to 1.34 g/ml. Centrifugation for 36 h and analysis of fractions containing the KBr density gradient ranging from 1.16 to 1.34 g/ml. Centrifugation for 36 h and analysis of fractions with densities of 1.20–1.25 g/ml were performed as described above. Pooled fractions were desalted and concentrated by ultrafiltration (Centriplus-100 filter units; Millipore, Eschborn, Germany) washing three times with TNE buffer. The concentration of the purified HBsAg was estimated by A<sub>280</sub>, assuming a value of 4.3 for 1 mg/ml (36), and by amino acid analysis (see below). Purified HBsAg was stored at −20 °C.

Isolation of tryptic L and M Protein-derived Peptides—Digestion of purified native HBsAg particles (4 mg) was carried out with trypsin (Sequencing Grade, Roche Molecular Biochemicals, Mannheim, Germany) in 1.2 ml of TNE buffer for 1 h at 37 °C using an enzyme: substrate ratio of 1:40 (w/w). Peptides and particles were separated by ultrafiltration using one Microcon-100 filter unit (Millipore). Tryptic peptides were isolated by reversed-phase HPLC (rHPLC) on a C<sub>18</sub>-column (5 μm, 30 mm, 4.6 × 250 mm; Vydac, Hesperia, CA) using 0.1% (v/v) aqueous trifluoroacetic acid with an acetonitrile gradient (0–60% in 60 min) and a flow rate of 1 ml/min at 37 °C. Peptides were monitored by absorption at 220 nm, and fractions were collected semi-automatically. Peptide-containing fractions were stored at −20 °C.

Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry (MALDI-TOF-MS) of Glycopeptides and Released Oligosaccharides—Molecular masses of rHPLC-purified peptides were determined by MALDI-TOF-MS on a Vision 2000 mass spectrometer (Finnigan MAT, Bremen, Germany). 1 μl of peptide solution (1–5 pmol) was mixed with 1 μl of matrix solution (10 mg of 2,5-dihydroxybenzoic acid/ml of 0.1% (v/v) trifluoroacetic acid, 30% (v/v) acetonitrile) and allowed to air-dry. Ions were generated by irradiation with a pulsed nitrogen laser (emission wavelength 337 nm; laser power density about 10<sup>6</sup> watts/cm<sup>2</sup>), and positive ions were accelerated and detected in the reflectron and linear mode. For analysis of released oligosaccharides, 50 pmol of glycans in aqueous solution were used and analyzed as described above. For calibration of the peptide mass spectra, human angiotensin and bovine insulin (both from Sigma, Deisenhofen, Germany) were used as external standards. For calibration of the mass spectra of released oligosaccharides, the diantennary oligosaccharide standard NA2 (Galβ4GlcNAcβ2Man3Galβ4GlcNAcβ2Man6βGlcNAc4GlcNAc; Oxford GlycoSciences, Abingdon, Oxfordshire, United Kingdom) and bovine insulin B (Sigma) were used. The accuracy of mass determination was about 0.03% for free oligosaccharides and (glyco)peptides.

Amino Acid Analysis—20–50 ng of protein or peptide were lyophilized and hydrolyzed in the gas phase over 6 h HCl, with 0.02% mercoptohanol, for 24 h at 110 °C. Free amino acids were dissolved in 0.1 mol of 0.5 N borate buffer, pH 7.7, derivatized with Fmoc-(N-(9-fluorenylmethoxycarbonyl), and analyzed by rHPLC on a Merck-Hitachi (Darmstadt, Germany) system composed of an AS-4000 autosampler, a L-6200A pump, a F-1050 fluorescence detector, and a D-6000 interface.

Peptide and Protein Sequencing—HPLC-purified peptides (50–100 pmol) or HBV subviral particles after tryptic digestion (about 500 pmol) were amino-terminally sequenced by automated Edman degradation on an Applied Biosystems (Foster City, CA) pulsed liquid phase sequencer, model 477A or 471A, under standard conditions. Phenyldihydrothiainodine-derivatives of amino acids were identified by an on-line analyzer, model 120A or 140B (Applied Biosystems), with a repetitive yield of 92–95%.

N-terminally blocked tryptic peptides were further digested with chymotrypsin (sequencing grade; Roche Molecular Biochemicals). Resulting products were fractionated by rHPLC and sequenced by Edman degradation as above.

Liberation of N-Linked Oligosaccharides from Glycopeptides—HPLC fractions containing the N-glycosylated pre-S2 tryptic peptides (20 μg, about 10 nmol) were pooled, lyophilized, resuspended in 150 μl of 20 mM sodium phosphate buffer, pH 7.5, and digested with 100 units of porcine pancreatic aminopeptidase M and 100 units of aminopeptidase N (N-acetyl-β-glucosaminidase) from Microbacterium meningosepticum (Roche Molecular Biochemicals) for 24 h at 37 °C. The digestion was repeated with 50 milliunits of enzyme. Oligosaccharides were isolated from peptide residues by rHPLC as described previously (37).

High pH Anion Exchange Chromatography (HPAEC) and Gel Filtration—Separation of released oligosaccharides was carried out at room temperature on a Dionex (Sunnyvale, CA) BioLC system using a CarboPac PA-100 column (4.6 × 250 mm) in series with a CarboPac PA guard column as described in detail earlier (38). A sodium acetate (210 mM) gradient (0–100% in 48 min) in 100 mM NaOH was used at a flow rate of 1 ml/min. Fractions (1 ml) were collected and immediately neutralized with 25 μl of 1 N acetic acid. Oligosaccharide-containing fractions were pooled, lyophilized, and resuspended in water. Desialylation was performed by Bio-Gel P2 (Bio-Rad) chromatography as reported earlier (39).

Carbohydrate Constituent and Methylation Analysis—The carbohydrate constituents were analyzed as detailed elsewhere (40). In short, hydrolysis of liberated oligosaccharides was performed with 4 M trifluoroacetic acid at 100 °C for 4 h; for hydrolysis of glycoproteins, 0.5 N sulfuric acid in 80% acetic acid (v/v) was used at 80 °C for 6 h. Following reduction with sodium borohydride and peracetylation, alditol acetates were analyzed by capillary gas-liquid chromatography/mass spectrom-
**Glycosylation of HBV Surface Proteins**

**TABLE I**

Characterization and assignment of tryptic peptides from HBsAg after separation by rHPLC

| Peak number | Retention time | Fraction | Peptide | Modifications | MALDI-TOF-MS<sup>a</sup> | Calculated Mass<sup>d</sup> | Number of sialic acids<sup>c</sup> |
|-------------|----------------|----------|---------|---------------|-------------------------|-----------------------------|-------------------------------|
| 1           | 33.2           | Ia       | Pre-S<sub>2</sub>(1–16) | Acetyl; 0, 1 or 2 O; diantennary N-glycan | 3672.2, 3688.9, 3689.0(2), 3693.0(1) | 2 or 1 | 2 or 1 |
| 2           | 33.5 + 33.8    | Ia       | Pre-S<sub>2</sub>(1–16) | Acetyl; 0 or 1 O; diantennary N-glycan | 3683.2, 3683.6 | 1 or 0 | 1 or 0 |
| 3           | 34.3           | Ia       | Pre-S<sub>2</sub>(1–16) | Acetyl; 1 O; diantennary N-glycan | 3672.0 | 1 or 0 | 1 or 0 |
| 4           | 35.1 + 35.4    | Ia       | Pre-S<sub>2</sub>(1–16) | 0 or 1 O; diantennary N-glycan | 3615.6, 3630.8, 3631.0(1) | 2, 1 or 0 | 2, 1 or 0 |
| 5           | 36.2           | Ia       | Pre-S<sub>2</sub>(1–16) | Acetyl; diantennary N-glycan | 3658.3 | 3657.0(0) | 2, 1 or 0 |
| 6           | 36.6           | Ia       | Pre-S<sub>2</sub>(1–16) | Acetyl; 0 or 1 O; diantennary N-glycan | 3656.2, 3672.5, 3673.0(1) | 2 or 1 | 2 or 1 |
| 7           | 36.9 + 37.1    | Ia       | Pre-S<sub>1</sub>(93–102) | | 1106.2 | 1106.3 | |
| 8           | 37.4 + 37.6    | Ia       | Pre-S<sub>2</sub>(6 to 16) | | 2655.4 | 2655.9 | 1 or 0 |
| 9           | 37.8 + 38.2    | Ia       | AlbumininT<sub>1</sub>38–180 | | 2531.4 | 2531.7 | |
| 10          | 38.4           | Ia       | Pre-S<sub>1</sub>(25–46) | | 2639.5 | 2639.9 | |
| 11          | 39.4           | Ib       | Pre-S<sub>2</sub>(6 to 16) | | 2639.5 | 2639.9 | |
| 12          | 40.5           | Ia       | Pre-S<sub>2</sub>(19–48) | GalNAc + Gal + | 3389.4, 3681.2 | 3681.1 | 1 |
| 13          | 41.3           | Ia       | Pre-S<sub>2</sub>(19–48) | GalNAc + Gal + | 3389.5 | 3389.8 | 0 |
| 14          | 42.8           | Ia       | Pre-S<sub>2</sub>(19–48) | GalNAc + Gal + | 3389.4, 3681.2 | 3681.1 | 1 |
| 15          | 43.7           | Ia       | Pre-S<sub>2</sub>(19–48) | GalNAc + Gal + | 3389.5 | 3389.8 | 0 |
| 16          | 44.6           | Ia       | Pre-S<sub>2</sub>(19–48) | GalNAc + Gal + | 3389.5 | 3389.8 | 0 |
| 17          | 45.5 + 45.8    | Ia       | Pre-S<sub>2</sub>(19–48) | GalNAc + Gal + | 3389.4, 3681.2 | 3681.1 | 1 |
| 18          | 47.5           | Ia       | Pre-S<sub>1</sub>(7–88) | GalNAc + Gal + | 3389.5 | 3389.8 | 0 |
| 19          | 48.7           | Ia       | Pre-S<sub>2</sub>(19–48) | GalNAc + Gal + | 3389.5 | 3389.8 | 0 |
| 20          | 49.8           | Ia       | Pre-S<sub>2</sub>(19–48) | GalNAc + Gal + | 3389.5 | 3389.8 | 0 |

**a** Amino acid residues derived from the pre-S<sub>2</sub> region are marked by underlined numbers.

**b** 0, 1, or 2 O, incorporation of zero, one or two oxygen atoms due to assumed oxidation of N-terminal Met.

**c** Pseudomolecular ions ([M+H]<sup>+</sup>) obtained in positive ion reflectron mode.

**d** Calculated masses are based on average masses. Numbers in parentheses (0, 1, or 2) indicate degree of Met oxidation.

**e** Deduced from pseudomolecular ions ([M+H]<sup>+</sup>) obtained by MALDI-TOF-MS in the linear mode.

**f** Including one sialic acid residue.

Peptides eluting in the time interval of 32–50 min (cf. Fig. 2) were identified by MALDI-TOF-MS, amino acid analysis, and N-terminal Edman sequencing. Peptides and assigned post-translational modifications are listed. The N-terminal pre-S<sub>2</sub> peptide (positions 1–16) was shown to be fully glycosylated and further modified, in part, by acetylation and/or oxidation. The corresponding peptide derived from the L protein (6 to 16) was also partially oxidized but not glycosylated at Asn-4. The tryptic peptide 19–48 of the pre-S<sub>2</sub> domain was found to be partially O-glycosylated. Peaks 12–14 were not assigned.

For linkage analyses, oligosaccharide alditols were permethylated using the instrumentation and microtechniques described earlier. For linkage analyses, oligosaccharide alditols were permethylated using the instrumentation and microtechniques described earlier.
digest were mixed with 1 μl of 2,5-dihydroxybenzoic acid on the target, and MALDI-TOF analysis was performed as described above.

**Quadrupole Time-of-flight (QTOF) Electrospray Ionization (ESI) Mass Spectrometry of Glycopeptides—** Positive ion ESI-MS and ESI-MS/MS was performed on a hybrid QTOF electrospray mass spectrometer (Micromass, Manchester, UK) in the nanospray mode. The glycopeptide ions were produced in an atmospheric pressure ionization/ESI ion source, using argon as a collision gas, and were transported to the mass spectrometer through a hexapole lens for optimal transmission. The nanospray capillaries were produced in the laboratory in Münster, using a Kopf vertical pipette puller (model 720, David Kopf Instruments, Tujunga, CA). The capillaries were not gold-coated, but an internal wire electrode was used.

Nanoelectrospray low energy collision-induced dissociation was performed as described recently (45) with the peptide 19–48 carrying one Gal-GalNAc disaccharide unit. The sample was dissolved in water to a concentration of 5 pmol/μl and introduced to the nanospray needle. The collision-induced dissociation conditions were optimized for the maximal signal intensity of glycosylated fragments (minimal deglycosylation).

For MS/MS, sequencing, the triply charged precursor ion was selected in the quadrupole, fragmented in the hexapole collision cell, refocused in the radio frequency-only-hexapole, and extracted orthogonally into the TOF analyzer.

**RESULTS**

**Isolation of the pre-S2 Glycopeptides—** HBV subviral particles were purified separately from the sera of two well-characterized HBV carriers infected with two different but typical strains of genotype D virus. In both cases, identical results were obtained. Therefore, only one set of data is presented.

To exclude contamination by glycopeptides containing Asn-146, native HBsAg particles were incubated with trypsin. Since the S domain is known to be highly resistant to trypsin, whereas pre-S cleavage sites are very sensitive (46), only glycopeptides of the pre-S region were released (Fig. 1). After removal of the digested HBsAg particles by ultrafiltration, released peptides were isolated by RP-HPLC (Fig. 2). Peptides eluting in the interval of 32–50 min were identified by amino acid analysis, MALDI-TOF-MS, and peptide sequencing (Table I). Compounds eluting prior to peptide 1 could be neither registered by MALDI-TOF-MS nor characterized by amino acid and sequence analysis and are, therefore, assumed to represent non-peptide contaminants. In parallel, residual trypsin-treated HBsAg particles (cf. Fig. 1B, lane 3) were also analyzed by Edman sequencing, demonstrating (a) complete removal of pre-S tryptic peptides, (b) Arg-48 to be the most C-terminal cleavage site accessible for trypsin in native particles, and (c) a molar ratio of about 20% of pre-S2-containing proteins (M and L proteins) versus 80% of S proteins in spherical particles.

The N-terminal pre-S2 tryptic peptide (1–16; cf. Fig. 1C) was found by MALDI-TOF-MS analysis to be N-terminally acetylated to an extent of about 90% and was, therefore, not directly accessible to Edman sequencing. Following further digestion of this glycopeptide with chymotrypsin and subfractionation of the resulting peptides by RP-HPLC, two internal fragments (9–12 and 13–16) were identified by their molecular masses, and a third (4–8) could be assigned by Edman degradation (data not shown). The original glycopeptide eluted from the RP-HPLC column as several successive peaks (marked as Ia in Fig. 2) due to its heterogeneity in glycosylation and acetylation and/or partial oxidation of the N-terminal Met yielding methionine sulfoxide or methionine sulfone derivatives. Oxidation of Met-1 is indicated by the observation that the original pre-S2 glycopeptide (1–16) displayed, in part, additional mass increments corresponding to one or two oxygen atoms, which were not registered in the case of chymotryptic subfragments (4–8, 9–12, and 13–16). Furthermore, removal of the N-terminal Met by BrCN cleavage was not possible, in agreement with literature data (47), which confirm that oxidized methionine residues are not cleaved by this treatment. It is not clear, however, whether the

**FIG. 3.** MALDI-TOF mass spectra of the products of sequential exoglycosidase digestion. A, native oligosaccharide fraction (about 50 pmol); B–D, products obtained after digestion with 0.4 milliunits of α-A. ureafaciens exoglycosidase digestion. B–D, 50 pmol); [M + 2H]2+ (1–16) displayed, in part, additional mass increments corresponding to one or two oxygen atoms, which were not registered in the case of chymotryptic subfragments (4–8, 9–12, and 13–16). Furthermore, removal of the N-terminal Met by BrCN cleavage was not possible, in agreement with literature data (47), which confirm that oxidized methionine residues are not cleaved by this treatment. It is not clear, however, whether the
isolated glycopeptides with peptide-saccharides, obtained after treatment of the HBV N-glycans. A, released oligosaccharides, obtained after treatment of isolated glycopeptides with peptide-N\(^\beta\)-asparagine amidase F, were fractionated by HPAEC using a CarboPac PA-100 column (4.6 \times 250 mm) and a gradient of sodium acetate in 0.1 M NaOH as indicated. Oligosaccharides were monitored by pulsed amperometric detection. B, structures proposed for the N-linked carbohydrate chains of the pre-S2 and S domains of HBsAg.

oxidation already occurred in vivo or reflects an artifact produced during sample handling in the laboratory. In accordance with literature data, we could further demonstrate that the M protein is N-glycosylated to 100% in the pre-S2 domain, whereas the L protein is not glycosylated in this position, as confirmed by the analysis of the peptide (–6 to 16; cf. Fig. 1, peptides 8 and 11 in Fig. 2, and Table I). By a similar line of evidence, the pre-S2 tryptic peptide (19–48) was found to be partially modified by O-glycosylation (see below).

**Structural Analysis of Pre-S2 N-Glycans**—For carbohydrate structure analysis, pre-S2 N-glycans were preparatively released from peptides 1–6 (cf. Fig. 2) by treatment with peptide-N\(^\beta\)-asparagine amidase F and isolated by rHPLC. Neutral carbohydrate constituent analysis revealed a molar ratio of GlcNAc, Man, and Gal of 3.4:3.0:2.2, which is typical for diantennary complex-type N-glycans. Sialic acid residues were not registered by the method employed.

This structure was verified by sequential exoglycosidase digestion in combination with MALDI-TOF-MS (Fig. 3) and by methylation analysis. MALDI-TOF-MS spectra of the released native oligosaccharide fraction displayed three major signals indicative of a complex-type diantennary N-glycan carrying zero, one, or two sialic acid residues (Fig. 3A). Sialylated oligosaccharides and glycopeptides are known to lose sialic acid very easily under MALDI conditions. Even with 6-aza-2-thiothymidine as matrix, usually less than 50% of sialylated molecules pass through the reflector intact (48). Therefore, it was not possible to quantify the degree of sialylation from these experiments. After treatment with sialidase from *A. ureafaciens*, peaks reflecting sialylated components shifted to the first one (Fig. 3B). Subsequent treatment with \(\beta\)-galactosidase from *D. pneumoniae*, cleaving only Gal(\(\beta\)1–4) linkages, caused a shift in the molecular mass approaching 324 Da, demonstrating the release of two hexose residues (Fig. 3C). Further digestion with \(\beta\)-N-acetylglucosaminidase from *D. pneumoniae*, known to split exclusively GlcNAc(\(\beta\)1–2) bonds, resulted in the loss of two GlcNAc residues and corresponding mass shifts of roughly 406 Da (Fig. 3D), yielding the molecular mass of the common pentasaccharide core of N-linked oligosaccharides. It is noteworthy that these experiments were performed with only 50 pmol of released N-glycans in a single spot directly on the MALDI target without intermediate purification or desalting procedures. The results were confirmed by methylation analysis of the total fraction of released N-glycans, which demonstrated the presence of 3,6-disubstituted Man, 2,3-substituted Man, 4,6-substituted GlcNAc, terminal Gal, and 6-substituted Gal in a molar ratio of about 0.8:2.0:1.3:0.3:1.3, respectively.

In order to determine the degree of sialylation, the released oligosaccharides were separated by HPAEC (Fig. 4A), which revealed that 50% of the glycans contained two sialic acids, 45% contained one, and about 5% contained no sialic acid substituent. The detailed structure of the carbohydrate chain is shown in Fig. 4B.

In order to compare pre-S2 N-glycans with the known structure of the N-glycan present in the S domain (49, 50), trypsin-treated subviral particles, obtained from sera of two different patients, from which the pre-S2 segments were completely removed (as confirmed by SDS-PAGE and Edman sequencing; see Fig. 1C, and see above), were subjected to carbohydrate constituent and methylation analysis. In both cases, the results obtained were compatible with the published diantennary complex-type structure carrying two, one, or zero terminal sialic acids (see Fig. 4B). In detail, Man, Gal, and GlcNAc were identified as neutral oligosaccharide components by carbohydrate constituent analysis in a molar ratio of 3.0:2.2:3.9. Furthermore, methylation analysis revealed the presence of 3,6-disubstituted Man, 2-substituted Man, 4-substituted GlcNAc, terminal Gal, and 6-substituted Gal in a molar ratio of about 1.3:2.0:1.5:0.3:1.6, respectively. The published low degree in fucosylation of about 5%, however, could not be confirmed by the methods used. Furthermore, we could exclude O-glycosylation of the S domain due to the absence of GalNAc.

**Identification and Structural Analysis of Pre-S2 O-Glycans**—The pre-S2 tryptic peptides (19–48), identified by amino acid analysis and Edman sequencing, were found to elute in four different peaks from the rHPLC column (Fig. 2, brackets IIa and IIb), each one of which displayed a different molecular mass in MALDI-TOF-MS (see Fig. 5 and Table I). The mass differences were consistent with the mass increments of monosaccharide components such as HexNAc, Hex, and NeuAc (cf. Table I). Carbohydrate constituent analysis of the glycopeptides (fraction IIa peptides in Fig. 2) confirmed the peptides to be modified by an O-glycosidically linked glycan, comprising GalNAc and Gal as neutral monosaccharide constituents. Sialic acid was not detectable by the procedure used. The extent of glycosylation was calculated from the peak areas in the rHPLC chromatogram (Fig. 2), resulting in about 40% to be not O-glycosylated, 5% to contain only one GalNAc-residue.
To elucidate the structure of the O-glycan, glycopeptides carrying the trisaccharide unit were treated sequentially with sialidase from *A. ureafaciens* and O-glycosidase from *D. pneumoic*, specifically releasing Gal(β1–3)GalNAcα-chains from Ser/Thr residues, directly on the MALDI target. Resulting products were analyzed by MALDI-TOF-MS. The results revealed mass shifts of about 293 and 365 Da, reflecting the release of one sialic acid and one Gal(β1–3)GalNAc unit, respectively, yielding the molecular mass of the unsubstituted peptide (data not shown; see also Ref. 43). The glycopeptide containing only the disaccharide unit was similarly susceptible to digestion with O-glycosidase.

In order to verify these conclusions, the O-glycopeptides were subjected to methylation analysis. The results supported the assignment, that the structure of the sialylated glycan is Neu5Ac(α2–3)Gal(β1–3)GalNAcα-Ser/Thr, i.e. the so called sialosyl-T-antigen (see Fig. 9B). The assumption that the reducing GalNAc-residue is α-glycosidically linked is based on the reported substrate specificity of the enzyme O-glycosidase, whereas the β-anomeric configuration of galactosyl residues was confirmed by on target digestion with β-galactosidase from bovine testes in combination with MALDI-TOF-MS (data not shown). Notably, sialic acid was found to be linked to C-3 of the terminal Gal, although *A. ureafaciens* sialidase is known to release preferentially (α2–6)-linked sialic acid residues. Possibly, the kinetic properties of this enzyme are influenced by the presence of matrix during on target digestion of the glycopeptide.

In order to allocate the observed O-glycosylation to M and/or L proteins, HBsAg filaments (*cf*. Fig. 1B, lane 4) were separated by SDS-PAGE and subjected to blotting and lectin analysis or immunostaining, before and after treatment with *A. ureafaciens* sialidase, using peanut agglutinin, specifically binding Gal(β1–3)GalNAc-units or specific monoclonal antibodies directed against M or L proteins. The results revealed both M protein subspecies to be O-glycosylated, whereas L proteins did not react with the lectin, demonstrating that this type of O-glycosylation is obviously a specific feature of the HBV M protein (Fig. 6).

Localization of the Pre-S2 O-Glycan—The sequence of the pre-S2 peptide (19–48) contains seven serine and three threonine residues. Using this peptide for prediction of the O-glycosylation site by the NetOGlyc 2.0 Prediction Server, which produces neural network predictions of mucin type GalNAc O-glycosylation sites in mammalian proteins (51), three threo-
nine (Thr-31, -37, and -38) and two serine residues (Ser-43 and -47) were predicted as potential O-glycosylation sites (see Fig. 1C), with Thr-37 having the highest potential of all. Similar results appeared by using the complete M or L protein sequences.

In order to chemically identify the site of O-glycosylation, glycopeptides 15, 16, and 17 as well as the nonglycosylated peptide 19 (see Fig. 2 and Table I) were digested individually with carboxypeptidases in combination with MALDI-TOF-MS analysis. Treatment of the unglycosylated peptide 19–48 (fraction 19; cf. Fig. 1C) led to a ladder of signals due to the removal of from one up to 18–20 amino acids.

In the case of the glycopeptides carrying the disaccharide or the trisaccharide, only up to 10 amino acids were removed, whereas the peptide modified with GalNAc only allowed the release of up to 11 amino acids (Fig. 7). These results pointed to Thr-37 being most likely the O-glycosylation site, since the carboxypeptidases used cleave only unmodified amino acids from the C terminus of peptides.

**Fig. 7.** MALDI-TOF mass spectra of O-glycosylated peptide 19–48 before and after digestion with carboxypeptidases. A, 15 pmol of native glycopeptide carrying one GalNAc residue (peak 17 in Fig. 2 and Table I); B, product obtained after extensive digestion (2 × 24 h) with carboxypeptidases P and Y from *P. janthinellum* and yeast. 11 amino acids are removed from the C terminus of the peptide, suggesting Thr-37 to be the linkage position of the glycan chain (cf. Fig. 1C). Conditions for MALDI-TOF-MS were as in Fig. 5.

Glycosylation of HBV Surface Proteins
To further substantiate this assignment, the glycopeptide carrying the Gal-GalNAc disaccharide unit was studied by QTOF analysis (Fig. 8). In the nanospray ESI-QTOF MS/MS experiment, both carbohydrate and peptide/glycopeptide sequence data were obtained in a single experiment. The triply charged molecular ion [M+3H]3+ at m/z = 1129.7 as a precursor ion. The fragment ions from the C-terminal y9 and N-terminal b19 ions are in accordance with the peptide sequence and assigned regarding their status of glycosylation. The linkage position of the O-glycan is deduced from the fragment ions at the two cleavage sites, b19 + HexNAc (found as the singly and as the doubly charged ion) as well as y13 + HexNAc (singly and doubly charged) and y13 + HexNAc + Hex (only singly charged) ions. These two glycopeptide fragment ions overlap at exactly one potential glycosylation site, i.e. at the Thr-37, providing direct evidence for glycan attachment. The Hex-HexNAc sugar chain structure is documented by the ion at m/z = 536.11 Da, besides the HexNAc diagnostic ion at m/z = 204.08. B, enlarged mass range details from the spectrum depicted in A: m/z = 810–1000 (top), m/z = 1550–1660 (middle), and m/z = 1820–2010 (bottom).
Glycosylation of HBV Surface Proteins

Fig. 8—continued

B

K. A. 10.06.98 nanoESI Q-TOF MSMS of m19-49 O-glyc. (1192.7 triple charged)

\[ y_{28}^+ \]

\[ y_{13} + \text{HexNAc}^{2+} \]

\[ y_{16}^+ \]

\[ y_{27}^+ \]

\[ y_{28}^+ \]

\[ y_{19}^+ \]

\[ y_{17} + \text{HexNAc} \]

\[ b_{17}^+ \]

\[ b_{30}^+ \text{HexNAc}^{2+} \]

\[ y_{15} \]

\[ y_{19}^+ \]

\[ y_{16} + \text{HexNAc} \]

\[ b_{20}^+ \]

\[ b_{21}^+ \]

\[ b_{19}^+ \text{HexNAc} \]

\[ b_{17} + \text{H}_2\text{O} \]

\[ y_{17} \]

\[ y_{14} + \text{HexNAc} \]

\[ y_{16} + \text{HexNAc} \]

\[ b_{20} \]

\[ -\text{H}_2\text{O} \]

\[ 1831.71 \]

\[ 1830.70 \]

\[ 1817.74 \]

\[ 1806.98 \]

\[ 1796.63 \]

\[ 1786.58 \]

\[ 1574.62 \]

\[ 1564.62 \]

\[ 1554.56 \]

\[ 1553.55 \]

\[ 1503.55 \]

\[ 1493.55 \]

\[ 1403.55 \]

\[ 1303.55 \]

\[ 1003.55 \]

\[ 1.0 \]

\[ 0.8 \]

\[ 0.6 \]

\[ 0.4 \]

\[ 0.2 \]

\[ 0.0 \]

m/z

%
Amino acid and sequence analyses, as well as MALDI-TOF-MS, of individual pre-S-derived tryptic and chymotryptic (glyco)peptides revealed that about 90% of the M protein is N-terminally acetylated. Possibly, acetylation of the aminoterminal end protects against proteolytic degradation. This may be particularly relevant for the pre-S2 domain, which is known to be highly sensitive toward proteolysis (46). This finding is remarkable in so far that it could be demonstrated that the recently identified M protein of WHV is not N-terminally blocked (9).

The third post-translational modification of the pre-S2 domain comprises O-glycosylation, which has similarly been found to be restricted to the M protein. Although there is evidence that HBV M protein expressed in mammalian cell culture (8) or WHV M protein (9) may be O-glycosylated, information on the number of carbohydrate substituents, their structures, and their linkage position(s) is still lacking. Hence, our results represent the first example for a detailed structural analysis of HBV M protein O-glycosylation. It is demonstrated that the M proteins under study carried, at least in part, a single O-linked carbohydrate substituent, which could be identified as GalNAcα- (Tn-antigen), Galβ(1–3)GalNAcα- (T-antigen), or Neu5Acα(2–3)Galβ(1–3)GalNAcα-unit (sialosyl T-antigen). Although neither the number nor the structures of the O-glycan(s) in M protein expressed in rodent cell culture (8) were investigated, the size increase of this protein, as evidenced by SDS-PAGE, was significantly larger than expected for the O-linked side chain described here. Likewise, the increase in size, due to O-glycosylation, was higher for WHV than for HBV M protein (9). Possibly, O-glycan substituents exist in rodent cell-derived M proteins in higher numbers and/or enlarged structures. Thus, the carbohydrate moieties, antigenicity and immunogenicity of the pre-S2 domains in certain candidate vaccines may differ from that of natural M proteins. For exact allocation of the O-glycosylation site, two different strategies were adopted: (a) exhaustive digestion of the glycopeptide with carboxypeptidases in conjunction with MALDI-TOF-MS and (b) ESI-QTOF-MS/MS analysis. Although the C-terminal half of the pre-S2 region is rich in Ser and Thr, both techniques revealed a specific substitution of the protein at Thr-37. This is consistent with recent findings (8) that M protein expressed in COS-7 cells is O-glycosylated between positions 27 and 47 of its amino acid sequence. Using the NetOGlyc 2.0 Prediction Server (51) for prediction of potential O-glycosylation sites, five Ser/Thr residues of the C-terminal pre-S2 peptide sequence are highlighted, with Thr-37 having the highest potential. It is interesting to note that in vivo only this amino acid is, in fact, glycosylated. Exact quantitative estimation of the degree of M protein O-glycosylation is not possible, since unglycosylated peptides 19–48 may be derived from both M and L proteins. Preliminary analyses of pre-S2-derived glyco/peptides from M protein, isolated by preparative SDS-PAGE and in situ trypsin digestion, similarly revealed the presence of unglycosylated peptides 19–48, ruling out a complete O-glycosylation of this protein (data not shown).

Since Thr-37, embedded in a sequence context favorable for O-glycosylation, is highly conserved in the pre-S2 domains of HBV genotypes B–F, in all primate HBV genomes including the recently described virus from the new world woolly monkey (52), and even in the otherwise highly divergent pre-S2 sequence of WHV, respective threonine residues might be similarly glycosylated, thus pointing to an as yet unknown function of potential evolutionary advantage. The two isolates of HBsAg analyzed in this study are typical but not totally identical representatives of HBV genotype D2 (53), which is prevalent in

---

4 S. Schaefer and W. H. Gerlich, unpublished observations.
Europe and the Middle East. The complete identity of the results, even in the proportions of the various glycan structures, suggests that the detected modifications of pre-S2 may be present in many, possibly most, HBV isolates. The O-glycosylation at Thr-37 of pre-S2, however, cannot be essential because genotype A of HBV has no Thr from position 32 to 55 of pre-S2 but contains numerous Ser residues in this region.

Glycoproteins with O-linked glycans have been found in a number of enveloped viruses including, for example, the M (E1) protein of murine corona virus, carrying exclusively O-glycans (54), and herpes simplex virus glycoproteins (55). Freind murine leukemia virus glycoprotein 71 (56), respiratory syncytial virus G protein (57), and Marburg virus glycoprotein (58), all of which contain both O- and N-glycans. Distinct functions of the O-linked glycans of viral glycoproteins are still obscure. It is speculated, however, that they may influence the antigenicity of the virion surface structure and protect the polypeptide against degradation.

One putative function of the pre-S2 O-glycan may reside in the partial masking of the respective peptide sequence. This assumption is corroborated by the finding that anti-pre-S2 antibodies are not readily detectable in recipients of pre-S2-containing vaccines and can be only transiently registered in patients with acute hepatitis B. In particular, mouse monoclonal antibodies directed against an epitope involving Thr-37 of pre-S2 may be masked by pre-S1 sequences.

The cytosolic orientation of the pre-S domains of the L protein allows alternative post-translational modifications such as N-glycosylation at Thr-37 of pre-S2, however, cannot be essential because the translocation of the pre-S(1 + 2) domain occurs in the medial to trans-Golgi or later. Alternatively, the Thr-37 of pre-S2 may be masked by pre-S1 sequences.

In agreement with earlier qualitative data on O-glycosylation (4), this study definitively proves that HBV L protein is not N-glycosylated in its pre-S region, thus supporting the assumption that the pre-S(1 + 2) domain of L protein is not translocated into the lumen of the endoplasmic reticulum during biosynthesis (13). The absence of O-glycans in L protein might suggest that the translocation of the pre-S(1 + 2) domain occurs in the medial to trans-Golgi or later. Alternatively, the Thr-37 of pre-S2 may be masked by pre-S1 sequences.

Microbial Infections (Cullier, L., Balows, A., and Sussman, M., eds) 9th Ed., pp. 745–747, Arnold, London

3. Heermann, K. H., Gerlich, W. H., Chudy, M., Schaefer, E., and Thomssen, R. (1999) J. Clin. Microbiol. 37, 68–73
4. Heermann, K. H., Goldmann, U., Schwartz, W., Seyfarth, T., Bauhammer, G., and Gerlich, W. H. (1984) J. Virol. 52, 396–402
5. Heermann, K. H., and Gerlich, W. H. (1991) in Molecular Biology of Hepatitis B Virus (McLachlan, A., ed) pp. 109–143, CRC Press, Boca Raton, FL

REFERENCES

1. Hollinger, F. B. (1995) in Fields Virology (Fields, B. N., Knipe, D. M., and Howley, P. M., eds) 3rd Ed., pp. 2739–2807, Lippincott-Raven, Philadelphia
2. Kann, M., and Gerlich, W. H. (1998) in Topley & Wilsons Microbiology and Molecular Biology of Hepatitis B Virus (McLachlan, A., ed) pp. 109–143, CRC Press, Boca Raton, FL

5. W. H. Gerlich, unpublished results.
6. D. Sobotta and W. H. Gerlich, unpublished results.
Brunak, S. (1998) *Glycoconj. J.* 15, 115–130

52. Lanford, R. E., Chavez, D., Brasky, K. M., Burns, R. B., III, and Rico-Hesse, R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 5757–5761

53. Bichko, V., Pushko, P., Dreilina, D., Pumpen, P., and Gren, E. (1985) *FEBS Lett.* 185, 208–212

54. Niemann, H., Geyer, R., Klenk, H. D., Linder, D., Stirn, S., and Wirth, M. (1984) *EMBO J.* 3, 665–670

55. Serafini-Cesi, F., Dall’Olio, F., Malagolini, N., Pereira, L., and Campadelli-Fiume, G. (1988) *J. Gen. Virol.* 69, 869–877

56. Geyer, R., Dabrowski, J., Dabrowski, U., Linder, D., Schütter, M., Schott, H. H., and Stirn, S. (1990) *Eur. J. Biochem.* 187, 95–110

57. Grauber, C., and Levine, S. (1985) *J. Gen. Virol.* 66, 417–432

58. Geyer, H., Will, C., Feldmann, H., Klenk, H. D., and Geyer, R. (1992) *Glycobiology* 2, 299–312

59. Persing, D. H., Varmus, H. E., and Ganem, D. (1987) *J. Virol.* 61, 1672–1677

60. Grbacic, E. V., and Anderson, D. A. (1994) *J. Virol.* 68, 7344–7350

61. Rothmann, K., Schnolzer, M., Radziwill, G., Hildt, E., Moelling, K., and Schaller, H. (1998) *J. Virol.* 72, 10138–10147

**Glycosylation of HBV Surface Proteins**

Glycosylation is a critical aspect of the biology of hepatitis B virus (HBV) surface proteins. It affects both the structural integrity and functional properties of these proteins, playing a key role in the virus’s ability to evade host immune responses. In this context, glycosylation patterns can significantly influence the immune response, viral replication, and pathogenesis. Understanding the glycosylation of HBV surface proteins is crucial for developing effective antiviral therapies. This section highlights recent findings in the field of HBV glycosylation, emphasizing the importance of this modification in the virus’s life cycle and its implications for disease management.