Journal of proteome research

Article

pubs.acs.org/jpr

Glycoproteomic Discovery of Serological Biomarker Candidates for HCV/HBV Infection-Associated Liver Fibrosis and Hepatocellular Carcinoma

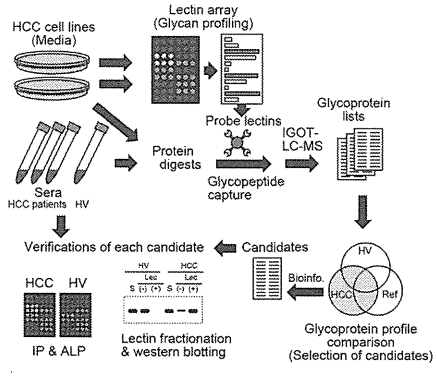
Hiroyuki Kaji,[†] Makoto Ocho,[†] Akira Togayachi,[†] Atsushi Kuno,[†] Maki Sogabe,[†] Takashi Ohkura,[†] Hirofumi Nozaki,[†] Takashi Angata,[†,¶] Yasunori Chiba,[†] Hidenori Ozaki,[†] Jun Hirabayashi,[†] Yasuhito Tanaka,[‡] Masashi Mizokami,[‡,§] Yuzuru Ikehara,[†] and Hisashi Narimatsu*[,†]

[†]Research Center for Medical Glycoscience, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan

[‡]Department of Virology and Liver Unit, Nagoya City University Graduate School of Medical Sciences, Mizuho, Nagoya, Aichi, Japan

S *Supporting Information*

ABSTRACT: We previously proposed a high-throughput strategy to discover serological biomarker candidates of cancer. This strategy focuses on a series of candidate glycoproteins that are specifically expressed in the original tissues (cells) of the target cancer and that carry glycan structures associated with carcinogenesis [Narimatsu, H., et al. *FEBS J.* **2010**, *277*(1), 95−105]. Here, we examined the effectiveness of our strategy in identifying biomarkers to assess progression of liver fibrosis and for the early detection of hepatocellular carcinoma (HCC). On the basis of the results of lectin array analyses in culture media of hepatoma cell lines, we captured glycopeptides carrying AAL-ligands (fucosylated glycans) or DSA-ligands (branched glycans) from digests of culture media proteins and sera from HCC patients with a background of liver cirrhosis (LC). Glycoproteins were identified by the IGOT-LC−MS method. In all, 21 candidates were selected from 744 AAL-bound glycoproteins for further verification according to (i) their abundance in serum, (ii) their specific expression in liver, and (iii) the availability of antibodies to the glycoproteins. All selected candidates showed enhancement of AAL-reactivity in sera of HCC patients compared with that of healthy volunteers (HV). These results indicate that our glycoproteomic strategy is effective for identifying multiple glyco-biomarker candidates in a high-throughput manner.



KEYWORDS: *glycoprotein, glycan alteration, hepatocellular carcinoma, liver fibrosis, biomarker, lectin array, glycoproteomics*

INTRODUCTION

In Japan, hepatocellular carcinoma (HCC) often develops in association with chronic liver diseases such as liver cirrhosis (LC) caused by persistent infection with hepatitis B or C virus followed by liver fibrogenesis.[1] The annual incidence of HCC increases with the degree of liver fibrosis and reaches 7−8% for the LC patients (fibrosis stage = F4).[2] Therefore, it is important to understand the degree of progression of liver fibrosis for patients with chronic hepatitis. This will enable identification of high-risk patients with progressed liver fibrosis, who should then undergo further detailed examinations such as computed tomography (CT) to detect evidence of early stage HCC. Presently, the progression of liver fibrosis is diagnosed by percutaneous biopsy; however, the method is both invasive and painful.[3,4] In addition, this method may result in sampling errors because the portion of the organ subjected to examination is very small.[5,6] These concerns prompted us to develop a serum biomarker to quantitatively evaluate the progression of liver fibrosis. Hence, we aimed to identify a marker that is applicable to a rapid, low cost, accurate, sensitive, and minimally invasive test. Furthermore, development of a serum HCC biomarker with increased sensitivity is required in order to diagnose early stage HCC at a size that is applicable for treatment such as radiofrequency ablation, thereby enhancing survival rates.[7,8]

Thus far, serological biomarkers have been screened by proteomic analyses using sera of HCC patients.[9,10] However, early stage cancer cells often express almost the same proteome as their originating cells because they have a common lineage. Hence, it is difficult to discover, verify, and validate practical serological marker candidates by comparing the levels of individual protein concentrations in serum between patients and healthy or disease controls. The problem can be seen by reference to a simple example. Suppose that the cancerous liver cells comprise 1% of the entire liver and secrete a 10-fold excess of a certain protein into the blood compared with the noncancerous cells. The blood concentration of such a protein in the cancer patient will be 1.1-fold that of the healthy individual.

Received: December 27, 2012
Published: April 15, 2013

If the secretion increases 100-fold, the blood concentration of the protein becomes approximately 2-fold; however, the difference is still less than that of individual variability. Alpha-fetoprotein (AFP), a clinical marker for HCC, is elevated from 3 ng/mL (i.e., average level in healthy individuals) to ≥20 ng/mL in a cancer patient,[11] which indicates HCC cells secrete >500-fold AFP by comparison to noncancerous cells. In fact, AFP was found to be a major component of the culture medium of a human hepatoma cell line (HepG2; data not shown). Thus, it is relatively straightforward to identify proteins that are significantly overexpressed by cancer cells (e.g., AFP). However, such proteins are not always found. These considerations reveal the difficulty of finding a "cancer-specific biomarker" based on the quantification of proteins.

We have established an integrated strategy for high-throughput discovery of serum biomarker candidates by utilizing a number of advanced technologies. Specifically, we developed techniques to analyze glycan biosynthesis-related genes (glycogenes), glycan structures, and glycoproteins as described previously.[12,13] We designed the strategy based on a well-known fact that the glycan structure on the cell surface and secretory proteins derived from cancerous cells is different from that derived from the same kind of noncancerous cells. Thus, we believe a qualitative alteration in the glycan structure, rather than a quantitative change in the amount of protein, will be a good index for developing serum biomarkers. The strategy comprises a 4-step analyses: (1) glycogene expression profiling by quantitative real-time PCR array for cancer cell lines, to predict possible and significant glycan structural motifs such as α1,6-fucose, bisecting GlcNAc (GlcNAcβ1-4Man), β1,6-branch-(GlcNAcβ1-6Man), poly-N-acetyllactosamine extension, Lewis antigen, and LacdiNAc (GalNAcβ1-4GlcNAc); (2) comparative glycan profiling by lectin array analysis of proteins secreted from cancer cells into the culture media, to determine cancer-associated glycan alteration and to select lectins to capture glycoproteins or glycopeptides displaying the altered glycans; (3) identification of core proteins carrying the cancer-associated glycans by LC–MS analysis; and (4) selection of promising glycoprotein candidates by comparative profiling of the identified glycoproteins in cancerous samples and healthy controls, followed by bioinformatics to establish tissue specificity and the expression level in the liver.

In accordance with this strategy, we digested glycoprotein mixtures prepared from the culture media of hepatoma cell lines, HepG2 and HuH-7, and also from the sera of HCC patients having a background of LC and those of healthy volunteers (HV). Glycopeptides were separately captured using AAL, DSA, and RCA120 lectins. Many glycoprotein biomarker candidates were identified (744 for AAL and 234 for DSA). In all, 21 AAL-bound glycoproteins were selected to determine the differences in their glycan structure between HCC patients and HV. All selected glycoproteins showed increases of AAL-reactivity in HCC sera, suggesting these would be valuable biomarkers of HCC, LC, or fibrosis progression. Our results confirm the effectiveness of this strategy in identifying cancer glycobiomarker candidates in a high-throughput manner.

METHODS

Materials

We collected sera from 8 HCC patients at Nagoya City University Hospital (NCUH) and control sera from 14 randomly selected healthy volunteers (HV) of 28–48 years of age at AIST, after obtaining the approval from the respective institutional ethics committees and the informed consent from all individuals for the use of clinical specimens. Clinical information of the HCC patients and information on sex and age of HV are presented in Supplementary Table 1. HV did not have hepatitis, other hepatic diseases, renal failure, clinical cardiovascular disease, or malignancies. Because we aimed at discovering not only biomarker candidates of HCC development but also specific markers for various stages between early phase fibrosis and severe liver cirrhosis, we used sera of HV, who were relatively younger than HCC patients, as control samples. To inactivate hepatitis viruses, each serum sample (both HV and HCC patients) was diluted 10-fold with Dulbecco's phosphate buffer containing 0.15 M NaCl (D-PBS) and 0.2% SDS and heated at 98 °C for 15 min before use. Human hepatoma cell lines (HepG2 and HuH-7) were obtained from RIKEN Cell Bank (Ibaraki, Japan). Details of all antibodies used in this study are presented in Supplementary Table 2.

Other reagents used in this study were as follows. Guanidine-HCl and neuraminidase (*Arthrobacter ureafaciens*, EC 3.2.1.18) were purchased from Nacalai Tesque (Kyoto, Japan). Dithiothreitol (DTT), iodoacetamide (IAA), trichloroacetic acid (TCA), methyl α-D-mannopyranoside, lactose, ethanol, 1-butanol, and trifluoroacetic acid (TFA) were from Wako Pure Chemicals (Osaka, Japan). Tosyl phenylalanyl chloromethyl ketone (TPCK)-treated trypsin was obtained from Thermo Fisher Scientific (Waltham, MA). Stable isotope-labeled water ($H_2{}^{18}O$, 99% atom% ^{18}O) was a product of Taiyo Nippon Sanso Corp. (Tokyo, Japan). L-Fucose and phenylmethylsulfonyl fluoride (PMSF) were obtained from Sigma-Aldrich (St Louis, MO). Glycopeptidase F (peptide-*N*-glycanase; PNGase) was from TakaraBio (Kyoto, Japan). The HPLC lectin column (LA-AAL, *Aleuria aurantia* lectin) and chitooligosaccharide were from Seikagaku Biobusiness (Tokyo, Japan). DSA (*Datura stramonium* agglutinin)-agarose and RCA (*Ricinus communis* agglutinin) 120-agarose were from Vector Laboratories, Inc. (Burlingame, CA). All solutions were prepared using ultrapure water. Analytical grade reagents were used unless stated otherwise.

HepG2 and HuH-7 were cultured in a Dulbecco's modified Eagle's medium (D-MEM; Life Technologies, Carlsbad, CA) supplemented with 10% heat-inactivated fetal bovine serum (FBS; JRH Biosciences, Lenexa, KS), 100 units/mL penicillin (Life Technologies), and 100 μg/mL streptomycin (Life Technologies) and maintained at 37 °C in an incubator with 5% CO_2. The cells at 60–80% confluence were washed 3 times with D-PBS and further cultured in FBS- and antibiotics-free D-MEM for 48 h. The final media were harvested, and the supernatants were filtered using a 0.45-μm disc filter (Millipore, Billerica, MA).

Preparation of Tryptic Digests of Culture Medium and Serum Proteins

The pooled HV serum was prepared by mixing equal volumes of heated sera of 14 HV. Proteins in the serum and culture media were collected by TCA precipitation (final concentration of 10%, w/w), S-reduced, and then alkylated as described previously.[14] For discovery experiments, we used 5 HCC sera (S-1 to S-5) listed in Supplementary Table 2. Proteins were digested with TPCK-treated trypsin, using an E/S ratio of 1/100 (w/w), at 37 °C for 16 h. Progression of the digestion was monitored by SDS-polyacrylamide gel electrophoresis followed by CBB staining. Protein concentrations were assayed by the BCA method using bovine serum albumin as a standard.

Lectin Affinity Capture of Glycopeptides from Tryptic Digests

Lectin affinity chromatography was performed as described previously.[14] Briefly, an aliquot of the tryptic digest was applied to an AAL column (4.6 mm i.d. × 150 mm) equilibrated with 10 mM HEPES-NaOH, pH 7.5. Glycopeptides were eluted with the same buffer containing 5 mM fucose. The pass-through fraction was applied again onto the same column, and residual glycopeptides were recovered using an identical procedure. Similarly, the trypsin digest was applied onto a DSA column (10 mm i.d. × 20 mm) after desialylation treatment with neuraminidase from *Arthrobacter ureafaciens*. Glycopeptides were then eluted with 1% chito-oligosaccharide (GlcNAc oligomers), and the eluate was directly applied onto an online RCA120 column (10 mm i.d. × 20 mm) to remove GlcNAc oligomers. Glycopeptides trapped on the RCA120 column were eluted with 0.1 M lactose. In addition, glycopeptides in the pass-through fractions of the AAL column chromatography were applied onto a RCA120 column (10 mm i.d. × 20 mm). Hereafter a fraction (protein/peptide) bound to a lectin column is denoted as "lectin(+)", e.g., AAL(+). Similarly, a passed (unbound) fraction is referred to as, e.g., DSA(−). Subsequently, all glycopeptide samples prepared (AAL(+), DSA(+), and AAL(−)/RCA120(+)) were further purified by hydrophilic interaction chromatography on a Sepharose CL-4B column as described previously.[14]

Enzymatic Stable Isotope Labeling of Glycopeptides

Purified glycopeptides were treated with PNGase (2.5 mU, 37 °C overnight) in stable isotope ^{18}O-labeled water to remove the glycan moiety and concomitantly label the glycosylated Asn of glycopeptides with the isotope, as described previously.[15] This step is referred to as isotope-coded glycosylation site-specific tagging (IGOT).

Identification of the Labeled Peptides by Nanoflow LC–MS Analysis

Stable isotope-labeled peptides were analyzed by the LC–MS method, as described previously.[16] Briefly, the peptide mixture was injected onto a C18 trap column (0.5 mm i.d. × 1 mm). After washing, the column was connected to a nanoflow LC system (flow rate: 100 nL/min), and the peptides were separated on a reverse phase (C18) tip column (150 μm i.d. × 70 mm) using a linear gradient of MeCN (0–35% in 0.1% formic acid) for 70 min. The eluted peptides were sprayed directly into a quadrupole time-of-flight hybrid mass spectrometer (Q-TOF Ultima; Waters-Micromass, Beverly, MA). The spectrometer was operated in data-dependent MS/MS mode.

MS/MS spectra were processed using MassLynx software (version 4.0, Micromass) to create peak list files with smoothing by the Savitzky-Golay method (window channels, ±3). The files were processed by the MASCOT algorithm (Version 2.1, Matrix Science) to assign peptides using the refseq protein sequence database (34,276 entries, downloaded on July 2007). The database search was performed by the MS/MS ion search method with the following parameters. Enzyme: Trypsin + Lys-C; Fixed modification: Carbamidomethyl (Cys); Variable modifications: Gln > pyro-Glu (N-term Gln), Oxidation (Met), Pyro-carbamidomethyl (N-term Cys), IGOT (Asn > Asp+^{18}O = +2.988261 Da, custom-made) (Asn); Peptide mass tolerance: 200 ppm; Fragment mass tolerance: 0.5 Da; Max missed cleavage: 2. All results from the peptide search were exported as CSV files and processed by Microsoft Excel. First, we selected the peptides with rank 1 and an expectation value <0.05. Then, we selected the peptides that contained one or more aspartic acids labeled with ^{18}O atoms (IGOT) at the position of Asn in the consensus sequence for *N*-glycosylation, Asn-Xaa-(Ser/Thr), where Xaa is any residue except Pro.

Immunoprecipitation and Lectin Array Analysis of the Glycoprotein Marker Candidates

Selected glycoproteins were enriched with their specific antibodies. Aliquots (10 μL each) of the pooled sera of HCC patients and HV were incubated overnight with 1 μg of biotinylated antibody at room temperature. The resultant complexes were trapped by 20 μL of streptavidin-conjugated magnetic beads (Dynabeads MyOne Streptavidin T1; Invitrogen) and then released by heating with 20 μL of 0.2% SDS in TBS. Glycan profiles of the released proteins were analyzed by the antibody-assisted lectin profiling (ALP) method.[17] Briefly, we analyzed array signals by using LecChip and GlycoStation Reader (GP Biosciences Ltd., Yokohama, Japan) according to a prescribed method.[17] The data were processed with the program software, Array-Pro Analyzer (Ver. 4.5, Media Cybernetics, Inc., Silver Spring, MD).

AAL Fractionation of Serum Proteins

The denatured serum (10 μL each) was incubated with 100 μL of AAL-agarose at 4 °C for 16 h with gentle shaking. After recovery of the supernatant (AAL(−) fraction), the gel was washed three times with PBSTx. Then, the gel was suspended with an AAL elution buffer (PBSTx containing 500 mM L-fucose) and shaken at 4 °C for 16 h. The supernatant (AAL(+) fraction) was then collected. Both AAL(+) and AAL(−) fractions were concentrated by ultrafiltration using an Amicon Ultra unit (Ultracel 3 kDa, Millipore Amicon) to adjust the volume of each fraction.

LCA-AAL Serial Lectin Fractionation of Serum Glycoproteins

The glycoproteins in the pooled sera of HCC patients and HV were separated by serial lectin fractionation using LCA- and AAL-conjugated gels (Kuno, A., et al. manuscript in preparation). Briefly, the 100-fold diluted serum (100 μL) was incubated with 100 μL of LCA-agarose at 4 °C for 5 h with gentle shaking, and the supernatant (LCA(−) fraction) was recovered. After washing, LCA-bound proteins were eluted with 100 μL of an LCA elution buffer (PBSTx containing 500 mM α-methyl-D-mannopyranoside). Then, the LCA(−) fraction was incubated with 50 μL of AAL-agarose at 4 °C for 16 h, and the supernatant (LCA(−)/AAL(−) fraction) was recovered. The LCA(−)/AAL(+) fraction was recovered with 500 mM L-fucose.

Western Blot Analysis

Protein samples including an equivalent volume of serum were subjected to 10% SDS-PAGE and transferred to a PVDF membrane (BioRad Laboratories, Hercules, CA). The membranes were blocked and incubated with each antibody. Finally, the treated membrane was reacted with the Western Lightning chemiluminescence reagent (Perkin-Elmer, Boston, MA) at room temperature and exposed to Hyperfilm ECL (GE Healthcare) for 60 or 300 s.

RESULTS

Selection of Lectins for Capturing Glycopeptides Carrying HCC-Related Glycans

A tissue-specific protein carrying a cancer-associated glycan is suitable as a biomarker for the cancer. To develop such glycoprotein biomarkers, we designed an integrated strategy as described previously.[12] In accordance with this strategy (Figure

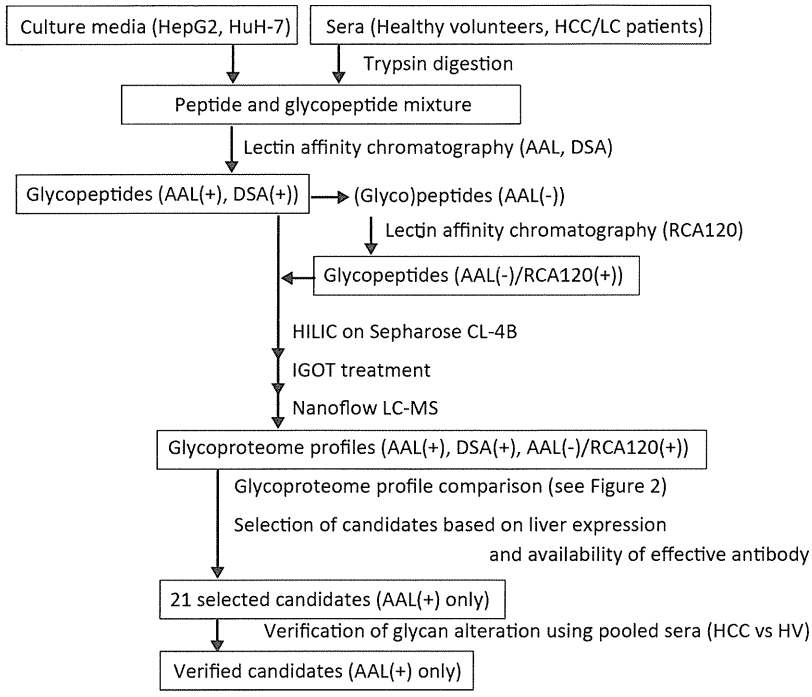


Figure 1. Experimental workflow of the discovery of serum glyco-biomarker candidates for liver disease. Glycopeptide fractions were captured with AAL- or DSA-affinity columns from tryptic digests of culture media (HepG2 and HuH-7) and serum taken from healthy volunteers and HCC patients. RCA-bound glycopeptides were recovered from pass fractions of the AAL-column for serum samples. After purification by HILIC, glycopeptides were identified by the IGOT-LC–MS method. From the AAL(+) glycoproteins, 21 glycoproteins were selected for further verification of HCC-associated glycan alteration (enhanced fucosylation). See Methods for further details.

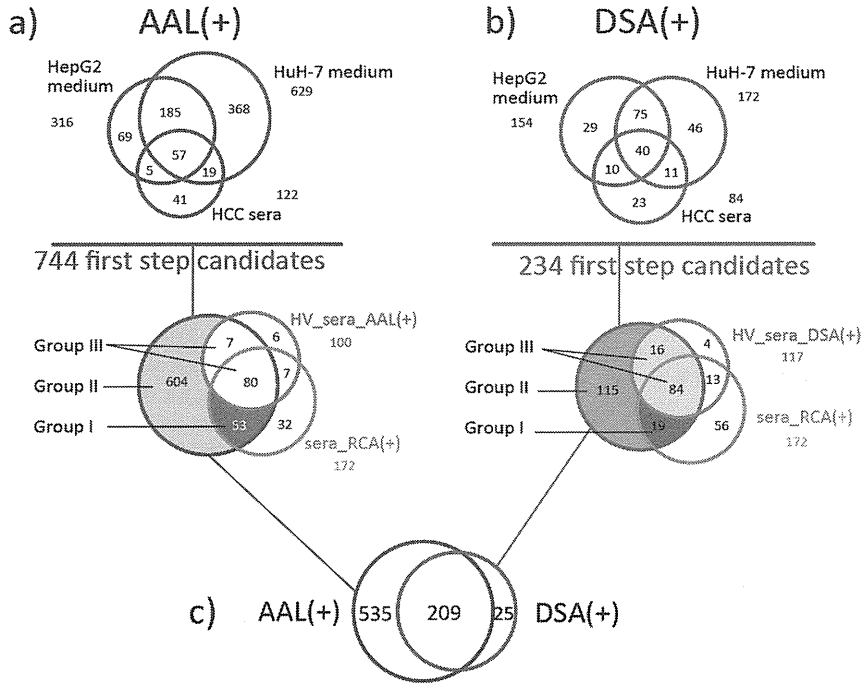


Figure 2. Venn diagram presentation for glycoprotein profile comparison. Lectin-bound glycopeptides were prepared from culture media (HepG2 and HuH-7) and sera of HCC patients and then identified by the IGOT-LC–MS method. In all, (a) 744 AAL(+) and (b) 234 DSA(+) glycoproteins were identified. AAL(+) and DSA(+) glycoproteins were identified from the sera of healthy volunteers (HV; red circle), and RCA120(+) glycoproteins were identified from the sera of HV and HCC patients (green circle). By comparing the glycoprotein profiles, HCC_AAL(+) proteins were classified into 3 groups. (c) Glycoproteins of HCC_AAL(+) and DSA(+) glycoproteins were compared. Most DSA(+) glycoproteins also possess fucosylated glycans.

1), we analyzed the glycan profiles of culture media of HCC cell lines (HepG2 and HuH-7)[13] and found that the lectin array signals of AAL and DSA as well as RCA120 were significant.

AAL has specificity for fucosylated glycans[18] and DSA for branched asialoglycans (tri- or tetra-antennary *N*-glycans).[19] It is known that the levels of these glycan motifs are increased by

dx.doi.org/10.1021/pr301217b | *J. Proteome Res.* 2013, 12, 2630–2640

Table 1. Biomarker Candidates Selected for Further Verification of HCC-Related Glycan Alteration (Fucosylation)[a]

				applicability of antibody		results of verification of the candidatens			
								serial lectin fractionation and WB	
candidate no. at verification step	group in Venn diagram (Figure 2)	gene symbols	description	WB (for sera)	IP (from sera)	IP and Lectin array (AAL)	AAL fractionation and WB	LCA(+)	LCA(−) AAL(+)
1	I	CPB2	plasma carboxypeptidase B2	√	√	ND	↑	→	→
2	I	CPN2	carboxypeptidase N, polypeptide 2, 83 kD	√	√	↑	NT	NT	NT
3	I	CSF1R	colony stimulating factor 1 receptor	√	√	↑	↑	↑	↑
4	I	FN1	fibronectin 1 isoform 7	√	−	NT	NT	→	↑
5	I	LAMP2	lysosomal-associated membrane protein 2	√	√	ND	NT	↑	→
6	I	SERPINA7	serine (or cysteine) proteinase inhibitor, clade A, member 7	√	−	NT	NT	→	↑
7	I	SHBG	sex hormone-binding globulin	√	−	NT	↑	↑	↑
8	II	ICAM2	intercellular adhesion molecule 2	√	−	NT	↑	→	→
9	II	SPARCL1	SPARC-like 1	√	√	→	↑	→	↑
10	III	AFM	afamin	√	−	NT	NT	↑	↑
11	III	AHSG	alpha-2-HS-glycoprotein	√	√	NT	NT	↑	→
12	III	APOD	apolipoprotein D	√	−	NT	NT	↑	→
13	III	AZGP1	alpha-2-glycoprotein 1, zinc	√	−	NT	NT	↑	↑
14	III	ICOSLG	inducible T-cell costimulator ligand	√	√	↑	↑	NT	NT
15	III	ORM1/2	orosomucoid 1/2 (= AGP)	√	√	↑[b]	NT	→	↑
16	III	PIGR	polymeric immunoglobulin receptor	√	√	↑	NT	↑	↑
17	III	PTGDS	prostaglandin H2 D-isomerase	√	−	NT	NT	↑	→
18	III	SEPP1	selenoprotein P isoform 1 precursor	√	−	NT	↑	↑	↑
19	III	SERPINA1	serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	√	√	NT	NT	→	↑
20	III	SERPINA3	serpin peptidase inhibitor, clade A, member 3	√	√	NT	NT	→	↑
21	III	TF	transferrin	√	√	NT	NT	↑	→

[a]Symbols: √, applicable; −, not applicable; NT, not tested; ND, not detected; ↑, enhanced in HCC; →, not enhanced in HCC. [b]See ref 32.

cancer-associated glycan alterations.[20−22] In addition, we showed that *N*-glycan fucosylation on AFP in the culture media was enhanced by comparison with that of placenta.[13] Therefore, we selected these lectins, AAL and DSA, to enrich HCC marker candidates. RCA120 recognizes Galβ1−4-GlcNAc-terminated glycans, which are present in the majority of serum glycoproteins after neuraminidase treatment.[23,24] Thus, RCA120 was used to comprehensively obtain glycopeptides from the samples in order to estimate their abundance.

Identification of Lectin-Captured Glycoproteins in the Culture Media and Serum Samples Using the IGOT-LC−MS Method

First, we collected glycopeptide subsets by AAL affinity chromatography from tryptic digests of the culture media of HepG2 and HuH-7. In order to evaluate the capturing specificity of the AAL column under the conditions employed, glycopeptides derived from the serum of an HCC patient were separated using the AAL column, and *N*-glycans released from AAL(+) and AAL(−) glycopeptides were analyzed by MALDI-TOF MS (Supplementary Figure 1 and Supplementary Method). As a result, almost all signals of AAL(+) glycans were assigned to have one or more deoxyhexose (dHex) residue(s) and only one slight signal having no dHex was observed, suggesting that the glycopeptides having fucosylated glycans were highly enriched in the AAL(+) fraction and that AAL has enough selectivity to capture glycopeptides having fucosylated glycans. However, because this result could not ensure that all individual glycopeptides were fucosylated, we dealt the AAL(+) glycopeptides as putative fucosylated glycopeptides. After purification by HILIC, the AAL(+) glycopeptides were treated with PNGase F in $H_2{}^{18}O$ and identified by LC−MS analysis followed by Mascot search. In total, 316 and 629 glycoproteins were identified from HepG2 and HuH-7, respectively (Figure 2a, Supplementary Tables 3 and 4). Similarly, we identified 122 AAL(+) glycoproteins from serum samples of 5 HCC patients. In total, 744 AAL(+) glycoproteins were identified as first-step biomarker candidates. Exemplary MS/MS spectra are shown in Supplementary Figure 2 to support the reliability of glycopeptide identification by the IGOT-LC−MS method. Next, we analyzed AAL(+) glycopeptides from HV sera and identified 100 glycoproteins. Finally, we used RCA120 to obtain glycopeptides from HV and HCC sera in order to estimate their abundance in serum and the number of *N*-glycosylated sites on each glycoprotein. Using this procedure, we identified 172 glycoproteins as abundant glycoproteins (Figure 2a).

The same experimental protocol allowed us to identify 154 and 172 DSA(+) glycoproteins from the culture media of HepG2 and HuH-7, respectively (Figure 2b). In addition, 84 DSA(+) glycoproteins were identified from HCC sera; in total, 234 glycoproteins were identified in HCC samples. Furthermore, 117 DSA(+) glycoproteins were identified from HV sera (Figure 2b). Details of the identified glycopeptides are given in Supplementary Table 3. Identified proteins in each sample are listed in Supplementary Tables 4 (AAL) and 5 (DSA). These results were registered to the glycoprotein database, GlycoProtDB, which was constructed using our experimental-based information and is

dx.doi.org/10.1021/pr301217b | J. Proteome Res. 2013, 12, 2630−2640

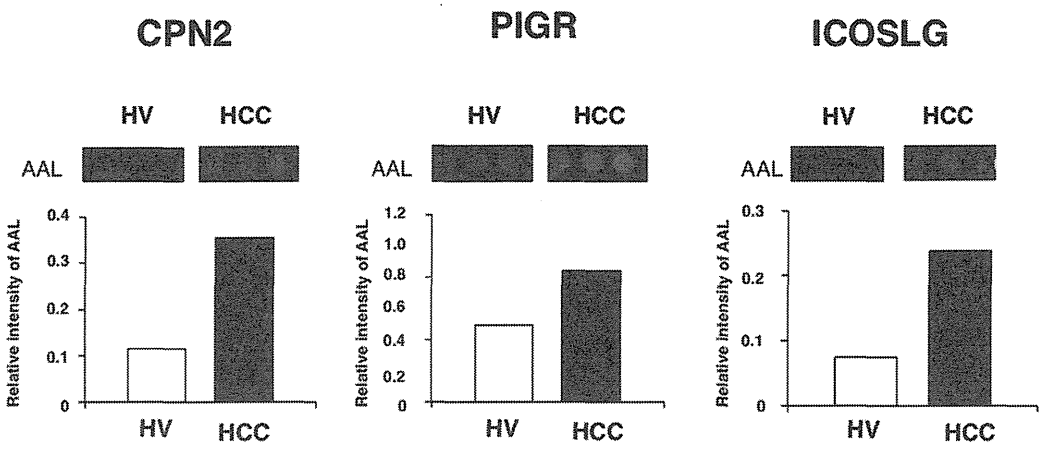


Figure 3. Detection of HCC-associated enhancement of fucosylation on the candidate glycoproteins by the immunoprecipitation-ALP method. Candidate glycoproteins were enriched from pooled sera of HV and HCC patients by immunoprecipitation with their antibodies and analyzed by lectin array (ALP method). Signal images for AAL spots (above, $n = 3$). Histogram indicating relative intensity of AAL to DSA (relative intensity of AAL signal against DSA signal).

currently available for mouse and *C. elegans* glycoproteins (http://jcggdb.jp/rcmg/gpdb/).[16] The overlap of AAL(+) and DSA(+) proteins was 209 (Figure 2c, Supplementary Table 6).

Selection of Biomarker Candidates for Further Verification by Glycoproteome Profiles and Availability of Their Antibodies

As described above, we identified 744 AAL(+) proteins in the media of HepG2 and HuH-7 and the serum samples of HCC patients. These proteins correspond to 398 gene symbols. Among them, 80 glycoproteins (gene symbols) are previously reported to have core-fucose in part or to bind with fucose-recognizing lectins.[25−28] The remaining 318 proteins are newly identified in this study as AAL-reactive proteins. Considering the selectivity of the AAL column under the employed conditions, as shown in Supplementary Figure 1, these are putative fucosylated glycoproteins that had not been formerly identified as fucosylated. Next, as a reference to select candidates to be verified, we classified the first-step AAL(+) candidates obtained from HCC samples into 3 groups by comparing the glycoproteome profiles as follows (Figure 2a). First, we compared these AAL(+) proteins with those in pooled HV serum (n = 100), and found 87 proteins in common. These common proteins were classified into Group III, whose *N*-glycans are also fucosylated in the healthy state. The other proteins were then compared with RCA120(+) serum proteins. This comparison identified 53 common proteins, suggesting that they are abundant in the serum. These proteins were classified into Group I. The remaining 604 proteins were designated as Group II. These proteins were detected only in the culture media, suggesting that they were less abundant in serum, but concentrated in the media. At this point, we minimized the number of proteins by using the gene symbols to avoid redundant identifications, such as splice variants and family proteins.

Next, we further selected the candidates on the basis of the tissue expression level and specificity for liver, as well as the availability of antibodies, which is essential to enrich or detect each candidate protein for verification. Moreover, glycoproteins supposedly produced by cells other than hepatocytes (e.g., immunoglobulins) were deleted from the list. We also eliminated some types of glycoproteins, such as complement factors and coagulation factors, as these were considered to be inappropriate for serological tests, because it is presumed easily that the level of complement components is influenced by inflammation in subject patients and the major parts of coagulation factors and fibrinogen are removed from serum samples during blood clotting (in preparation of serum from blood). Consequently, a total of 21 glycoproteins (gene symbols) were selected for further verification (Table 1). These candidates included glycoproteins known to have fucosylated glycans.[29,30] Nonetheless, we verified the glycan alteration between HCC and HV to confirm the effectiveness of our strategy.

The DSA(+) glycoproteins were also classified into 3 groups (Figure 2b) in the same manner as for the AAL(+) proteins. However, we did not select these candidates for further verification for the following reasons. First, although Group I (19) included glypican 3 (GPC3), which was reported to be a useful tumor marker of HCC and some other cancers,[31] other proteins were mostly coagulation factors, complement components, and fibrinogens, which are deemed unsuitable as described above. Thus, we could not find any suitable candidates in Group I of DSA(+) except GPC3. The DSA(+) candidates belonging to Group II (115) were identified only in the culture media, which means low abundance in serum. Furthermore, many of the DSA(+) proteins in Group III (100) were the same as those selected from the AAL(+) proteins. A further technical difficulty was that DSA(+) proteins were barely released from DSA-agarose in the preliminary experiments. Thus, subsequent verification of HCC-associated glycan alteration was performed only for the AAL(+) candidates.

Verification of HCC/LC-Associated Glycan Alteration (Enhanced Fucosylation) on the Candidate Glycoproteins

We verified HCC/LC-associated glycan alteration of the selected candidates using pooled sera of HV and HCC patients (with LC as background) by one or more of the following methods: (1) immunoprecipitation followed by lectin array analysis, (2) AAL-protein fractionation followed by Western blotting, and (3) LCA/AAL-serial lectin fractionation followed by Western blotting. Because these analyses require appropriate antibody for each protein, we tested commercially available antibodies using a pooled serum of HV (data not shown). Antibodies applicable for our verification analyses are listed in Supplementary Table 2.

dx.doi.org/10.1021/pr301217b | J. Proteome Res. 2013, 12, 2630−2640

1. Immunoprecipitation Followed by Lectin Array Analysis. Enhancement of fucosylation associated with HCC was confirmed by purification of the candidate proteins using immunoprecipitation followed by antibody-assisted lectin profiling (ALP) analysis[17] Among the 8 candidate proteins analyzed by this method, carboxypeptidase N polypeptide 2 (CPN2), polymeric immunoglobulin receptor (PIGR) and inducible T-cell costimulator ligand (ICOSLG) showed an apparent increase of AAL signal in the HCC sera compared to HV. As shown in Figure 3, the AAL signal of HCC sera was increased compared with the respective HV sera by the following factors: 3.1-fold for CPN2, 1.7-fold for PIGR, and 4.1-fold for ICOSLG. Results of the remaining candidates are shown in Supplementary Figure 3a and summarized in Table 1. The number of candidate proteins applicable to the ALP analysis is limited due to the lack of suitable antibodies. However, the ALP analysis also provided the reactivity information against other lectins, which is useful for reselecting the most appropriate lectin(s) for detection of glycan alteration associated with HCC/LC (fibrosis) as exemplified by α1-acid glycoprotein (AGP or ORM1/2).[32]

2. AAL-Fractionation Followed by Western Blotting Analysis. We fractionated pooled sera of HV and HCC patients into AAL(+) and AAL(−) fractions using an AAL column. To estimate the amount of candidates in each fraction, the protein fractions and untreated sera were subjected to SDS-PAGE, and then each candidate was visualized by Western blotting. The levels of all 7 candidates tested by this analysis increased in the AAL(+) fraction of HCC sera compared with HV, suggesting enhanced fucosylation associated with HCC/LC. Specifically, as shown in Figure 4, sex hormone binding globulin (SHBG: Group I), colony stimulating factor 1 receptor (CSF1R: Group I), intracellular adhesion molecule 2 (ICAM2: Group II), and selenoprotein P (SEPP1: Group III) indicated significantly increased signals in AAL(+) of HCC sera compared with that of HV sera. Results of other 3 candidates are shown in Supplementary Figure 3b.

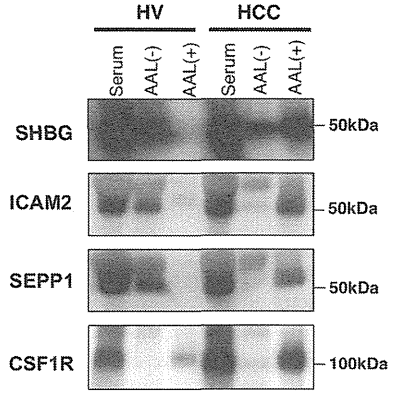


Figure 4. Detection of HCC-associated enhancement of fucosylation on the candidate glycoproteins by AAL-fractionation followed by Western blotting. Pooled sera of HV and HCC patients were fractionated with AAL-agarose. AAL(+), AAL(−), and untreated sera were separated by SDS-PAGE. The candidate glycoproteins were detected by Western blotting. Blot images of each candidate are shown.

3. LCA-AAL Serial Lectin Fractionation Followed by Western Blotting Analysis. As described above, branching and fucosylation of glycans are known to be enhanced in association with HCC.[20,22] Thus, measuring the glycoproteins that underwent both fucosylation and branching might be a clearer indication of glycan alteration by HCC compared to estimating the overall fucosylation level. In many cases, fucosylated glycoproteins are enriched by LCA or AAL; however, their specificities are different. LCA binds specifically to α1,6-fucosylated biantennary glycans, and AAL binds to fucosylated glycans regardless of branched structures.[33] Therefore, we separated the fucosylated candidate glycoproteins into 2 fractions by LCA-AAL serial lectin fractionation (refer to schematic shown in Figure 5a). First, a serum sample was applied to the LCA column to capture glycoproteins having α1,6-fucosylated biantennary glycans, and then the pass fraction (LCA(−)) was applied to the AAL column to recover the remaining glycoproteins having fucosylated glycans. Thus, by serial fractionation glycoproteins having both fucosylated and branched glycans are enriched in the LCA(−)/AAL(+) fraction. These fractions were separated by SDS-PAGE, and then each candidate protein was detected by Western blotting (Figure 5b). As a result, ORM1/2 (AGP), which is known to have highly branched glycans, was detected only in the LCA(−)/AAL(+) fraction (data not shown), indicating that the serial fractionation worked as intended. As summarized in Table 1, the ratio of the LCA(−)/AAL(+) fraction was higher in the HCC sera than HV sera for 12 of the 19 candidates tested (Figure 5b for SHBG, PIGR, CSF1R, and SEPP1; Supplementary Figure 3c for the remaining). These observations suggest that HCC/LC-associated increases of fucosylation and branching occur on the common glycans. In total, 16 of the 21 candidates were identified in both AAL(+) and DSA(+) fractions. Hence, glycans on a wide range of glycoproteins display both fucosylation and branching. However, the bands of proteins in the LCA(+) fraction also increased in HCC sera, suggesting that fucosylation associated with HCC/LC is independent of branching. Conversely, there were also proteins whose associated glycan was less highly branched or not branched at all. For 6 candidates, only the rate of the LCA(+) fraction was higher in the HCC sera (e.g., LAMP2 and SPARCL1). Thus, serial lectin fractionation may be useful for confirming the onset phase of fucosylation and branching individually in the course of chronic fibrosis, LC to HCC.

DISCUSSION

Currently, AFP, PIVKA-II (protein induced by vitamin K absence or antagonists-II, also known as des-γ-carboxy prothrombin [DCP]) and AFP-L3% (ratio of LCA-reactive AFP to total AFP) have been used for the diagnosis of HCC. However, the sensitivity and specificity of these markers are insufficient for diagnosis of early stage HCC.[34] Although many candidate serum biomarkers for HCC have been reported by proteomics- and/or glycoproteomics-based research, there is still a clinical need to develop more reliable markers for HCC.

However, it is fundamentally difficult to discover a serum biomarker for early stage cancer by comparing serum protein profiles between patients and healthy volunteers or disease controls. Empirically, we know that glycan alteration of particular cells is a good index for detecting a change of cell status such as differentiation and carcinogenesis. However, it is also difficult to detect early stage cancer by comparative profiling of the entire glycome of serum glycoproteins, i.e., by analyzing glycans released from serum glycoproteins *in vitro*. This is because the altered glycans emerge "ectopically" in association with carcinogenesis and thus do not have structures "specific to the cancerous cells". For example, sialyl Lewis X is known as a cancer-related antigen in colon,[35] non-small cell lung,[36] breast,[21] and pancreatic

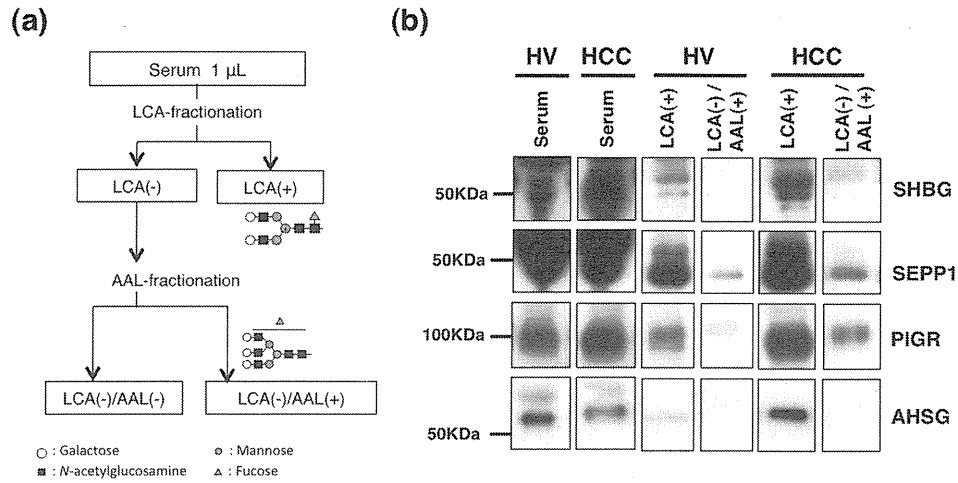


Figure 5. Detection of HCC-associated enhancement of fucosylation on the candidate glycoproteins by LCA/AAL-serial lectin chromatography followed by Western blotting. (a) Outline of the serial fractionation. Serum was fractionated with LCA-agarose. Glycoproteins in the LCA(+) fraction were presumed to have core-fucosylated biantennary glycans. Then, LCA(−) proteins were fractionated with AAL-agarose. Glycans of AAL(+) glycoproteins were presumed to be both fucosylated and branched. (b) Lectin-bound fractions were separated by SDS-PAGE, and then the candidate proteins in each fraction were detected by Western blotting.

cancers.[37] However, the glycan structure is also expressed by noncancerous cells such as endothelial cells of high endothelial venules (HEV) and lymphocytes associated with inflammation.[38] Similarly, although enhanced fucosylation is recognized as a glycan alteration related to HCC, fucosylation does not necessarily occur only in HCC, and fucosylated glycans are expressed ubiquitously. Therefore, it is important to detect glycan alteration on hepatic cell-specific proteins, especially when seeking a serum marker. In order to clarify glycan alteration associated with HCC, we first analyzed glycomes of the "culture media" of HCC cell lines (HepG2 and HuH-7) by lectin array as described previously.[13] Proteins in the culture media are undoubtedly secreted by HCC cells and carry glycans synthesized in the cancer cells. Thus, we used the culture media for the glycome analysis. In this sense, it is important to use a cell line that maintains *in vivo* characteristics of the cancer for the discovery procedure. In HCC, enhanced fucosylation and branching were reported in both cell lines and cancer tissues.[20,22] As expected, we found that the array signals of AAL and DSA in the media were very strong. Therefore, we selected the two lectins, AAL and DSA, to enrich glycoproteins carrying their cognate glycans, i.e., fucosylated and branched glycans, respectively. As exemplified by AFP-L3, which is AFP bearing α1,6-(core) fucosylated *N*-glycan, enhancement of fucosylation in HCC is a promising marker for cancer-associated glycan alterations.[39,40] Indeed, AFP-L3 is used as a U.S. Food and Drug Administration-approved HCC serum marker.

Neutral loss-triggered MS/MS (Electron Transfer Dissociation or Collision-Induced Dissociation) identification after partial deglycosylation with Endo F3 is a novel technology for assigning core fucose-bearing glycopeptides without employing lectin affinity. This method has been developed and used for marker discovery of HCC.[28,41] In general, however, glycoproteins having fucosylated *N*-glycans were enriched using fucose-recognizing lectins such as AAL, LCA, PSA, and VFA to capture not only the core fucose-bearing glycoproteins/peptides but also terminal α1,2/1,3-fucose-bearing molecules.[42]

Next, we identified AAL(+) and DSA(+) glycopeptides from the culture media of HCC cell lines to ensure they were actually produced by liver cancer cells. In total, 703 AAL(+) and 211 DSA(+) glycoproteins were identified from the media (Figure 2). These represent currently the largest data sets of human glycoproteins having a particular glycan motif (fucosylation or branch). It is well recognized that proteins obtained by affinity chromatography include those bound to the column nonspecifically or indirectly. For this reason, we captured glycopeptides, rather than glycoproteins, to minimize nonspecific identification. The IGOT method employed in this study also served to distinguish glycopeptides from non-*N*-glycosylated peptides with high precision.[15]

In addition, we also identified AAL(+) and DSA(+) glycopeptides from HCC sera. As the majority of serum glycoproteins are known to be derived from liver, serum samples are appropriate only for the marker discovery of hepatic diseases. We identified 703 and 122 AAL(+) glycoproteins from the culture media and the patients' sera, respectively. Roughly two-thirds of the AAL(+) glycoproteins from HCC sera (81/122) were also detected in the media, suggesting these glycoproteins were derived from liver cells. The remaining AAL(+) glycoproteins (41) were identified solely from the HCC sera, some of which are known to be primarily expressed in the liver. Thus, we selected 5 glycoproteins as candidates for further verification from the serum AAL(+) proteins.

Next, we compared the glycoproteome profiles as presented in Figure 2 to classify the candidates into three groups based on the following two factors. Due to the reasons outlined above, we confine our discussion to AAL(+) proteins only. The first factor is the degree of fucosylation under a physiological state. A preferable cancer marker is a glycoprotein having little or no fucosylated glycans in the healthy state. Thus, we divided the AAL(+) candidate proteins according to whether they were identified only in the HCC samples or in common with HV. Glycoproteins identified in common were classified into Group III (Figure 2). Another factor is the abundance of the core protein in serum. The RCA120(+) proteins identified from serum samples were thought to be highly abundant in serum. In addition, the number of RCA(+) peptides identified from each protein indicates the number of glycosylation sites on the target glycan. The more glycans the protein has, the easier the protein can be detected, for example, by sandwich ELISA using

dx.doi.org/10.1021/pr301217b | J. Proteome Res. 2013, 12, 2630−2640

antibodies and lectins, because multivalent binding strengthens the affinity between glycans on the glycoprotein and lectin molecules. The AAL(+) proteins identified in HCC samples that were in common with RCA120(+) proteins were classified into Group I. The remaining proteins were designated into Group II as shown in Figure 2. From each group, we selected 21 candidates for further verification based on the availability of antibodies and their abundance in serum. These candidates were then verified by one or more of the subsequent methods (Table 1).

Among 7 candidates in Group I, 4 proteins were already known to have enhanced fucosylation in HCC by comparison to the healthy state.[25−28] Glycoproteins of Group I include AFP, GP73 (GOLM1), and GPC3. Fucosylated AFP, known as AFP-L3, is actually being used in clinical practice for diagnosis of HCC. Moreover, fucosylated GP73 was reported as a marker for HCC with higher sensitivity than AFP.[26,43] These results suggest our approach for identifying candidate markers has been successful. In this study, all selected candidates in Group I (including CPB2, CSF1R, LAMP2, and SHBG) were newly identified AAL(+) glycoproteins displaying enhanced levels of fucosylation (Table 1). Among them, we expected CSF1R and CPN2 to be promising candidate HCC markers. Specifically, in comparison with other candidates, CSF1R and CPN2 show lower fucosylation in HV and increased expression in the liver and have many potential *N*-glycosylation sites (7 in CPN2;[44] 9 in CSF1R[45]), which would facilitate detection by sandwich ELISA using their respective antibodies and AAL lectin.

Glycoproteins in Group III included many classical plasma proteins as represented by transferrin, hemopexin, alpha 2-macroglobulin, kininogen, ceruloplasmin, complement factors, and so on.[46] Among 11 selected candidates, all but ICOSLG were reported to show elevated fucosylation in HCC.[25] However, these proteins in Group III were suggested to have fucosylated glycans in healthy individuals. This also suggests that these Group III proteins might be biomarkers that reflect the status of the liver as a whole rather than a local change, i.e., carcinogenesis. Namely, if the ratio of fucosylation on a glycan of a candidate biomarker is 1% in the healthy state and enhanced to 100% in HCC, the amount of this protein having fucosylated glycan would increase only 2-fold at most in a patient with a small cancer, i.e., 1% by weight (1*100:1*99 + 100 = 1:2). Thus, as discussed above, these proteins may become surrogate fibrosis markers. These markers are also clinically desirable, because a quantitative index indicating the degree of fibrosis is important in identifying high-risk individuals, who should be referred for imaging and cytological diagnosis (biopsy), e.g., hepatitis patients infected by HBV or HCV. Moreover, such an approach will be useful for evaluating the benefit of treatment or medication in patients and for the development of new medicines for liver fibrosis.[47] In this sense, ORM1/2 (AGP), PIGR, SERPINA3, AZGP1, ICOSLG, PTGDS, and SEPP1 might be good markers for fibrosis. In fact, we found that glycan alteration on ORM1/2 correlated with progression of liver fibrosis, which can diagnose LC with an accuracy of over 90%.[5,48,49]

In Group II, the ratio of AAL(+) of 2 candidates, SPARCL1 and ICAM2, increased in HCC. Elevated expression of SPARCL1 is known to be associated with the progression of hepatic disease along with sinusoidal capillary transformation.[50] As the serum level of these candidates is low, improvement of detection sensitivity will be needed for practical use as biomarkers.

As demonstrated in this study, we could identify many glycoproteins carrying glycans with enhanced reactivity to AAL in HCC sera and hepatoma cell lines using our high-throughput strategy. Furthermore, this approach also enabled us to provide a large set of data on human glycoproteins having particular glycan motifs through our glycoprotein database, GlycoProtDB, http://jcggdb.jp/rcmg/gpdb/.

In this study, we also carried out a verification of glycan alteration between HCC patients and HV for many candidate proteins. The verification procedure involved three methods including the newly developed technique of serial lectin fractionation. Because we used sera of HCC patients having LC as samples for the discovery and verification, the glycan alteration is thought to be associated with HCC or LC. Therefore, for each marker candidate, it should be verified at which stage on the long-term pathological course toward HCC the enhancement of glycan alteration (fucosylation) occurs. Detailed comparison using sera of patients at various stages of the disease may allow us to identify stage-specific biomarkers. We have already verified that glycan alteration on ORM1/2 (AGP in Group III) can distinguish the fibrosis stage between F3 and F4 (LC) by immunopurification followed by a lectin array analysis using sera from over 100 patients.[32] Furthermore, subsequent validation studies of this marker using an automated bedside clinical chemistry analyzer in more than 1000 HBV/HCV-infected patients have demonstrated that the diagnostic accuracy for detecting LC is superior to that of other fibrosis indices or instruments.[5,48,49] Thus, we expect that a novel and clinically applicable glycobiomarker for early stage HCC can be selected using the stage-established samples in a future study. Furthermore, we believe that this approach is applicable to discover biomarkers not only for other cancers but also for fibrogenic diseases such as Alzheimer's disease and fibrosis of lung and kidney.

■ ASSOCIATED CONTENT

Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +81-29-861-3200. Fax: +81-29-861-3201. E-mail: h.narimatsu@aist.go.jp.

Present Addresses

[¶]Institute of Biological Chemistry, Academia Sinica, 128 Academia Road, Sec. 2, Nankang, Taipei 11529, Taiwan.

[§]The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Chiba 272−8516, Japan.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by a grant of "Medical Glycomics Project" from New Energy and Industrial Technology Development Organization (NEDO). We thank Dr. Hiroaki Tateno, Ms. Jinko Murakami and Yoshiko Kubo for preparation of the lectin array; Ms. Sachiko Unno and Kozue Saitou for lectin array analyses; Ms. Nana Tanaka and Yoshie Takahashi and Mr. Norihito Hayatsu for technical assistance; Ms. Naoko Tetsura for

data processing of mass spectrometry, and Dr. Mika Fujita and Ms. Azusa Tomioka for MS analyses of glycans. We appreciate Drs. Takashi Sato, Hiromichi Sawaki, Masanori Gotoh and Hiromi Ito and Mr. Toshihide Shikanai for valuable discussion. We also thank Ms. Azumi Takahashi for assistance in the preparation of this manuscript. All persons are members of our research center.

REFERENCES

(1) Umemura, T.; Ichijo, T.; Yoshizawa, K.; Tanaka, E.; Kiyosawa, K. Epidemiology of hepatocellular carcinoma in Japan. *J. Gastroenterol.* **2009**, *44* (Suppl 19), 102–7.

(2) Yoshida, H.; Shiratori, Y.; Moriyama, M.; Arakawa, Y.; Ide, T.; Sata, M.; Inoue, O.; Yano, M.; Tanaka, M.; Fujiyama, S.; Nishiguchi, S.; Kuroki, T.; Imazeki, F.; Yokosuka, O.; Kinoyama, S.; Yamada, G.; Omata, M. Interferon therapy reduces the risk for hepatocellular carcinoma: national surveillance program of cirrhotic and noncirrhotic patients with chronic hepatitis C in Japan. IHIT Study Group. Inhibition of Hepatocarcinogenesis by Interferon Therapy. *Ann. Intern. Med.* **1999**, *131* (3), 174–81.

(3) Cadranel, J. F.; Rufat, P.; Degos, F. Practices of liver biopsy in France: results of a prospective nationwide survey. For the Group of Epidemiology of the French Association for the Study of the Liver (AFEF). *Hepatology* **2000**, *32* (3), 477–81.

(4) Castera, L.; Negre, I.; Samii, K.; Buffet, C. Pain experienced during percutaneous liver biopsy. *Hepatology* **1999**, *30* (6), 1529–30.

(5) Ito, K.; Kuno, A.; Ikehara, Y.; Sugiyama, M.; Saito, H.; Aoki, Y.; Matsui, T.; Imamura, M.; Korenaga, M.; Murata, K.; Masaki, N.; Tanaka, Y.; Hige, S.; Izumi, N.; Kurosaki, M.; Nishiguchi, S.; Sakamoto, M.; Kage, M.; Narimatsu, H.; Mizokami, M. LecT-hepa, a glyco-marker derived from multiple lectins, as a predictor of liver fibrosis in chronic hepatitis C patients. *Hepatology* **2012**, *56* (4), 1448–56.

(6) Wai, C. T.; Greenson, J. K.; Fontana, R. J.; Kalbfleisch, J. D.; Marrero, J. A.; Conjeevaram, H. S.; Lok, A. S. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* **2003**, *38* (2), 518–26.

(7) Sato, M.; Tateishi, R.; Yasunaga, H.; Horiguchi, H.; Yoshida, H.; Matsuda, S.; Koike, K., Mortality and morbidity of hepatectomy, radiofrequency ablation, and embolization for hepatocellular carcinoma: a national survey of 54,145 patients. *J. Gastroenterol.* **2012**.

(8) Minami, Y.; Kudo, M. Radiofrequency ablation of hepatocellular carcinoma: a literature review. *Int. J. Hepatol.* **2011**, 104685.

(9) Takayama, T.; Makuuchi, M.; Hirohashi, S.; Sakamoto, M.; Yamamoto, J.; Shimada, K.; Kosuge, T.; Okada, S.; Takayasu, K.; Yamasaki, S. Early hepatocellular carcinoma as an entity with a high rate of surgical cure. *Hepatology* **1998**, *28* (5), 1241–6.

(10) Rifai, N.; Gillette, M. A.; Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **2006**, *24* (8), 971–83.

(11) Tamura, Y.; Igarashi, M.; Kawai, H.; Suda, T.; Satomura, S.; Aoyagi, Y. Clinical advantage of highly sensitive on-chip immunoassay for fucosylated fraction of alpha-fetoprotein in patients with hepatocellular carcinoma. *Dig. Dis. Sci.* **2010**, *55* (12), 3576–83.

(12) Narimatsu, H.; Sawaki, H.; Kuno, A.; Kaji, H.; Ito, H.; Ikehara, Y. A strategy for discovery of cancer glyco-biomarkers in serum using newly developed technologies for glycoproteomics. *FEBS J.* **2010**, *277* (1), 95–105.

(13) Ito, H.; Kuno, A.; Sawaki, H.; Sogabe, M.; Ozaki, H.; Tanaka, Y.; Mizokami, M.; Shoda, J.; Angata, T.; Sato, T.; Hirabayashi, J.; Ikehara, Y.; Narimatsu, H. Strategy for glycoproteomics: identification of glyco-alteration using multiple glycan profiling tools. *J. Proteome Res.* **2009**, *8* (3), 1358–67.

(14) Kaji, H.; Yamauchi, Y.; Takahashi, N.; Isobe, T. Mass spectrometric identification of N-linked glycopeptides using lectin-mediated affinity capture and glycosylation site-specific stable isotope tagging. *Nat. Protoc.* **2006**, *1* (6), 3019–27.

(15) Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.* **2003**, *21* (6), 667–72.

(16) Kaji, H.; Shikanai, T.; Sasaki-Sawa, A.; Wen, H.; Fujita, M.; Suzuki, Y.; Sugahara, D.; Sawaki, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Takahashi, N.; Isobe, T.; Narimatsu, H. Large-scale identification of N-glycosylated proteins of mouse tissues and construction of a glycoprotein database, GlycoProtDB. *J. Proteome Res.* **2012**, *11* (9), 4553–66.

(17) Kuno, A.; Kato, Y.; Matsuda, A.; Kaneko, M. K.; Ito, H.; Amano, K.; Chiba, Y.; Narimatsu, H.; Hirabayashi, J. Focused differential glycan analysis with the platform antibody-assisted lectin profiling for glycan-related biomarker verification. *Mol. Cell. Proteomics* **2009**, *8* (1), 99–108.

(18) Matsumura, K.; Higashida, K.; Ishida, H.; Hata, Y.; Yamamoto, K.; Shigeta, M.; Mizuno-Horikawa, Y.; Wang, X.; Miyoshi, E.; Gu, J.; Taniguchi, N. Carbohydrate binding specificity of a fucose-specific lectin from Aspergillus oryzae: a novel probe for core fucose. *J. Biol. Chem.* **2007**, *282* (21), 15700–8.

(19) Green, E. D.; Brodbeck, R. M.; Baenziger, J. U. Lectin affinity high-performance liquid chromatography: interactions of N-glycanase-released oligosaccharides with leukoagglutinating phytohemagglutinin, concanavalin A, Datura stramonium agglutinin, and Vicia villosa agglutinin. *Anal. Biochem.* **1987**, *167* (1), 62–75.

(20) Moriwaki, K.; Miyoshi, E. Fucosylation and gastrointestinal cancer. *World J. Hepatol.* **2010**, *2* (4), 151–61.

(21) Listinsky, J. J.; Siegal, G. P.; Listinsky, C. M. The emerging importance of alpha-L-fucose in human breast cancer: a review. *Am. J. Transl. Res.* **2011**, *3* (4), 292–322.

(22) Guo, H. B.; Zhang, Y.; Chen, H. L. Relationship between metastasis-associated phenotypes and N-glycan structure of surface glycoproteins in human hepatocarcinoma cells. *J. Cancer Res. Clin. Oncol.* **2001**, *127* (4), 231–6.

(23) Shinohara, Y.; Sota, H.; Kim, F.; Shimizu, M.; Gotoh, M.; Tosu, M.; Hasegawa, Y. Use of a biosensor based on surface plasmon resonance and biotinyl glycans for analysis of sugar binding specificities of lectins. *J. Biochem.* **1995**, *117* (5), 1076–82.

(24) Itakura, Y.; Nakamura-Tsuruta, S.; Kominami, J.; Sharon, N.; Kasai, K.; Hirabayashi, J. Systematic comparison of oligosaccharide specificity of Ricinus communis agglutinin I and Erythrina lectins: a search by frontal affinity chromatography. *J. Biochem.* **2007**, *142* (4), 459–69.

(25) Comunale, M. A.; Wang, M.; Hafner, J.; Krakover, J.; Rodemich, L.; Kopenhaver, B.; Long, R. E.; Junaidi, O.; Bisceglie, A. M.; Block, T. M.; Mehta, A. S. Identification and development of fucosylated glycoproteins as biomarkers of primary hepatocellular carcinoma. *J. Proteome Res.* **2009**, *8* (2), 595–602.

(26) Drake, R. R.; Schwegler, E. E.; Malik, G.; Diaz, J.; Block, T.; Mehta, A.; Semmes, O. J. Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. *Mol. Cell. Proteomics* **2006**, *5* (10), 1957–67.

(27) Jia, W.; Lu, Z.; Fu, Y.; Wang, H. P.; Wang, L. H.; Chi, H.; Yuan, Z. F.; Zheng, Z. B.; Song, L. N.; Han, H. H.; Liang, Y. M.; Wang, J. L.; Cai, Y.; Zhang, Y. K.; Deng, Y. L.; Ying, W. T.; He, S. M.; Qian, X. H. A strategy for precise and large scale identification of core fucosylated glycoproteins. *Mol. Cell. Proteomics* **2009**, *8* (5), 913–23.

(28) Chen, R.; Wang, F.; Tan, Y.; Sun, Z.; Song, C.; Ye, M.; Wang, H.; Zou, H. Development of a combined chemical and enzymatic approach for the mass spectrometric identification and quantification of aberrant N-glycosylation. *J. Proteomics* **2012**, *75* (5), 1666–74.

(29) Hadziyannis, E.; Sialevris, K.; Georgiou, A.; Koskinas, J. Analysis of serum alpha-fetoprotein-L3% and des-gamma carboxyprothrombin markers in cases with misleading hepatocellular carcinoma total alpha-fetoprotein levels. *Oncol. Rep.* **2012**, No. 23174906.

(30) Biondi, A.; Malaguarnera, G.; Vacante, M.; Berretta, M.; D'Agata, V.; Malaguarnera, M.; Basile, F.; Drago, F.; Bertino, G. Elevated serum levels of Chromogranin A in hepatocellular carcinoma. *BMC Surg.* **2012**, *12* (Suppl 1), S7.

(31) Mounajjed, T.; Zhang, L.; Wu, T. T. Glypican-3 expression in gastrointestinal and pancreatic epithelial neoplasms. *Hum. Pathol.* **2012**, *44* (4), 542–50.

(32) Kuno, A.; Ikehara, Y.; Tanaka, Y.; Angata, T.; Unno, S.; Sogabe, M.; Ozaki, H.; Ito, K.; Hirabayashi, J.; Mizokami, M.; Narimatsu, H. Multilectin assay for detecting fibrosis-specific glyco-alteration by means of lectin microarray. *Clin. Chem.* **2011**, *57* (1), 48–56.
(33) Tateno, H.; Nakamura-Tsuruta, S.; Hirabayashi, J. Comparative analysis of core-fucose-binding lectins from Lens culinaris and Pisum sativum using frontal affinity chromatography. *Glycobiology* **2009**, *19* (5), 527–36.
(34) Masuzaki, R.; Karp, S. J.; Omata, M. New serum markers of hepatocellular carcinoma. *Semin. Oncol.* **2012**, *39* (4), 434–9.
(35) Yusa, A.; Miyazaki, K.; Kimura, N.; Izawa, M.; Kannagi, R. Epigenetic silencing of the sulfate transporter gene DTDST induces sialyl Lewisx expression and accelerates proliferation of colon cancer cells. *Cancer Res.* **2010**, *70* (10), 4064–73.
(36) Mizuguchi, S.; Inoue, K.; Iwata, T.; Nishida, T.; Izumi, N.; Tsukioka, T.; Nishiyama, N.; Uenishi, T.; Suehiro, S. High serum concentrations of Sialyl Lewisx predict multilevel N2 disease in non-small-cell lung cancer. *Ann. Surg. Oncol.* **2006**, *13* (7), 1010–8.
(37) Peracaula, R.; Tabares, G.; Lopez-Ferrer, A.; Brossmer, R.; de Bolos, C.; de Llorens, R. Role of sialyltransferases involved in the biosynthesis of Lewis antigens in human pancreatic tumour cells. *Glycoconjugate J.* **2005**, *22* (3), 135–44.
(38) Lee, J. K.; Bistrup, A.; van Zante, A.; Rosen, S. D. Activities and expression pattern of the carbohydrate sulfotransferase GlcNAc6ST-3 (I-GlcNAc6ST): functional implications. *Glycobiology* **2003**, *13* (4), 245–54.
(39) Taketa, K.; Ichikawa, E.; Sakuda, H.; Iwamasa, T.; Hayakawa, M.; Taga, H.; Hirai, H. Lectin reactivity of alpha-fetoprotein in a case of renal cell carcinoma. *Tumour Biol,* **1989**, *10* (5), 275–80.
(40) Medical devices; immunology and microbiology devices; classification of AFP-L3% immunological test systems. Final rule. *Fed. Regist.* **2005**, *70* (191), 57748–50.
(41) Zhao, Y.; Jia, W.; Wang, J.; Ying, W.; Zhang, Y.; Qian, X. Fragmentation and site-specific quantification of core fucosylated glycoprotein by multiple reaction monitoring-mass spectrometry. *Anal. Chem.* **2011**, *83* (22), 8802–9.
(42) Comunale, M. A.; Lowman, M.; Long, R. E.; Krakover, J.; Philip, R.; Seeholzer, S.; Evans, A. A.; Hann, H. W.; Block, T. M.; Mehta, A. S. Proteomic analysis of serum associated fucosylated glycoproteins in the development of primary hepatocellular carcinoma. *J. Proteome Res.* **2006**, *5* (2), 308–15.
(43) Marrero, J. A.; Feng, Z.; Wang, Y.; Nguyen, M. H.; Befeler, A. S.; Roberts, L. R.; Reddy, K. R.; Harnois, D.; Llovet, J. M.; Normolle, D.; Dalhgren, J.; Chia, D.; Lok, A. S.; Wagner, P. D.; Srivastava, S.; Schwartz, M. Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology* **2009**, *137* (1), 110–8.
(44) Tan, F.; Weerasinghe, D. K.; Skidgel, R. A.; Tamei, H.; Kaul, R. K.; Roninson, I. B.; Schilling, J. W.; Erdos, E. G. The deduced protein sequence of the human carboxypeptidase N high molecular weight subunit reveals the presence of leucine-rich tandem repeats. *J. Biol. Chem.* **1990**, *265* (1), 13–9.
(45) Wang, Z. E.; Myles, G. M.; Brandt, C. S.; Lioubin, M. N.; Rohrschneider, L. Identification of the ligand-binding regions in the macrophage colony-stimulating factor receptor extracellular domain. *Mol. Cell. Biol.* **1993**, *13* (9), 5348–59.
(46) Mitchell, P. Proteomics retrenches. *Nat. Biotechnol.* **2010**, *28* (7), 665–70.
(47) Rosenberg, W. M.; Voelker, M.; Thiel, R.; Becka, M.; Burt, A.; Schuppan, D.; Hubscher, S.; Roskams, T.; Pinzani, M.; Arthur, M. J. Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology* **2004**, *127* (6), 1704–13.
(48) Kuno, A.; Ikehara, Y.; Tanaka, Y.; Saito, K.; Ito, K.; Tsuruno, C.; Nagai, S.; Takahama, Y.; Mizokami, M.; Hirabayashi, J.; Narimatsu, H. LecT-Hepa: A triplex lectin-antibody sandwich immunoassay for estimating the progression dynamics of liver fibrosis assisted by a bedside clinical chemistry analyzer and an automated pretreatment machine. *Clin. Chim. Acta* **2011**, *412* (19–20), 1767–72.
(49) Du, D.; Zhu, X.; Kuno, A.; Matsuda, A.; Tsuruno, C.; Yu, D.; Zhang, Y.; Ikehara, Y.; Tanaka, Y.; Zhang, X.; Narimatsu, H. Comparison of LecT-Hepa and FibroScan for assessment of liver fibrosis in hepatitis B virus infected patients with different ALT levels. *Clin. Chim. Acta* **2012**, *413* (21–22), 1796–9.
(50) Lau, C. P.; Poon, R. T.; Cheung, S. T.; Yu, W. C.; Fan, S. T. SPARC and Hevin expression correlate with tumour angiogenesis in hepatocellular carcinoma. *J. Pathol.* **2006**, *210* (4), 459–68.

UICC global cancer control

IJC
International Journal of Cancer

Novel point mutations and mutational complexes in the enhancer II, core promoter and precore regions of hepatitis B virus genotype D1 associated with hepatocellular carcinoma in Saudi Arabia

Anis Khan[1,2], Mohammed A. Al Balwi [1,3,4*], Yasuhito Tanaka[2], Ali Hajeer[3,4], Faisal M. Sanai[5], Ibrahim Al Abdulkarim[1], Latifah Al Ayyar[1], Motasim Badri[4], Dib Saudi[6], Waleed Tamimi[3], Masashi Mizokami[2,7] and Bandar Al Knawy[8]

[1] King Abdullah International Medical Research Center, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia
[2] Department of Virology and Liver Unit, Graduate School of Medical Sciences, Nagoya City University, Nagoya, Japan
[3] Department of Pathology and Laboratory Medicine, King Abdulaziz Medical City, Riyadh, Saudi Arabia
[4] College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia
[5] Department of Hepatobiliary Sciences and Liver Transplantation, King Abdulaziz Medical City, Riyadh, Saudi Arabia
[6] Department of Gastroenterology, Military Hospital, Khamis Mushait, Saudi Arabia
[7] Research Center for Hepatitis and Immunology, International Medical Center Konodai Hospital, Ichikawa, Japan
[8] Department of Gastroenterology, King Abdulaziz Medical City, Riyadh, Saudi Arabia

In this study, a cohort of 182 patients [55 hepatocellular carcinoma (HCC) and 127 non-HCC] infected with hepatitis B virus (HBV) in Saudi Arabia was investigated to study the relationship between sequence variation in the enhancer II (EnhII), basal core promoter (BCP) and precore regions of HBV genotype D (HBV/D) and the risk of HCC. HBV genotypes were determined by sequencing analysis and/or enzyme-linked immunosorbent assay. Variations in the EnhII, BCP and precore regions were compared between 107 non-HCC and 45 HCC patients infected with HBV/D, followed by age-matched analysis of 40 cases *versus* equal number of controls. Age and male gender were significantly associated with HCC ($p = 0.0001$ and $p = 0.03$, respectively). Serological markers such as aspartate aminotransferase, albumin and anti-HBe were significantly associated with HCC ($p = 0.0001$ for all), whereas HBeAg positivity was associated with non-HCC ($p = 0.0001$). The most prevalent HBV genotype was HBV/D (94%), followed by HBV/E (4%), HBV/A (1.6%) and HBV/C (0.5%). For HBV/D1, genomic mutations associated with HCC were T1673/G1679, G1727, C1741, C1761, A1757/T1764/G1766, T1773, T1773/G1775 and C1909. Age- and gender-adjusted stepwise logistic regression analysis indicated that mutations G1727 [odds ratio (OR) = 18.3; 95% confidence interval (CI) = 2.8–118.4; $p = 0.002$], A1757/T1764/G1766 (OR = 4.7; 95% CI = 1.3–17.2; $p = 0.01$) and T1773 (OR = 14.06; 95% CI = 2.3–84.8; $p = 0.004$) are independent predictors of HCC development. These results implicate novel individual and combination patterns of mutations in the *X*/precore region of HBV/D1 as predictors of HCC. Risk stratification based on these mutation complexes would be useful in determining high-risk patients and improving diagnostic and treatment strategies for HBV/D1.

Key words: hepatitis B virus, sequence analysis, genotype D, *X*-gene, point mutations, mutational complexes, hepatocellular carcinoma
Grant sponsor: King Abdullah International Medical Research Center; **Grant number:** RR/010/01; **Grant sponsors:** Ministry of Health, Labor, and Welfare of Japan, Ministry of Education, Culture, Sports, Science, and Technology of Japan
DOI: 10.1002/ijc.28307
History: Received 22 Apr 2013; Accepted 14 May 2013; Online 6 Jun 2013
Correspondence to: Mohammed A. Al Balwi, Department of Pathology and Laboratory Medicine, King Abdullah International Medical Research Center, King Abdulaziz Medical City, P.O. Box 22490, Riyadh 11426, Kingdom of Saudi Arabia, Tel.: +966-1-252-0088, Fax: +966-1-252-0130, E-mail: balwim@ngha.med.sa

Approximately three billion people in the world are exposed to hepatitis B virus (HBV), of whom 350–400 million are persistently infected with it.[1] The two primary clinical manifestations associated with chronicity of the disease are cirrhosis and hepatocellular carcinoma (HCC), either of which may lead to liver-related death. It is estimated that chronic HBV carriers have a 100-fold increased risk for developing HCC compared to noncarriers[2]; however, the incidence rates of developing HCC vary widely around the world.

HBV has been characterized into eight genotypes (A–H) based on a divergence over the entire genome of greater than 8%.[3] Two new genotypes, I and J, have recently been reported and await international recognition.[4,5] These HBV genotypes are known to have a distinct geographical distribution. Most studies on clinical outcome in relation to HBV

What's new?

The accumulation of mutations in hepatitis B virus (HBV) over the course of long-term infection may increase its carcinogenicity, leading to the development of hepatocellular carcinoma (HCC). The combined effect of multiple mutations, however, has not been explored in detail. Here, novel mutations in the BCP and precore regions of HBV subgenotype D1 were associated with HCC in a study population in Saudi Arabia. The associations held for individual mutations and for combination patterns involving multiple mutations. The mutation complexes may help identify patients at high risk for HCC and could influence treatment strategies for HBV/D1.

genotypes have their origin in East Asia, having been conducted on patients infected with genotypes B and C. Genotype C is generally considered to be more virulent than genotype B.[6,7] Reports are emerging that subgenotypes within a genotype may also differ in the capacity to induce HCC based on viral sequence variations and recombination.[8]

In addition to the HBV genotype, variations in the *X*/basal core promoter (BCP)/precore regions have been shown to affect viral functions *in vitro*. Many functional sequences of HBV genome, such as enhancer II (EnhII), BCP, X-terminal signal, start points of two pregenomic messenger RNA, poly A signal and Epsilon, lie within this region.[9] Nucleotide changes in this region are therefore presumed to have a high carcinogenic capability.[10–12] Most of the previous studies have focused on the evaluation of individual mutations or only the combined effect of BCP double mutation T1762/1764 with respect to the development of HCC rather than evaluating the combined effect of multiple mutations or pattern of combinations.[13,14] As mutations accumulate gradually during long-term HBV infection, it is better to evaluate the combined effect of multiple mutations, which is expected to impart greater stress on the liver leading to HCC.

HBV genotype D (HBV/D) is the most prevalent genotype in South and Central Asia and the Middle East. To date, Six subgenotypes of HBV/D (D1–D6) have been identified and are distributed throughout the world.[15,16] The predominance of each subgenotype differs geographically and as such its role in the natural history of HBV infection may differ. It has been reported that genotype D is the most prevalent genotype in Saudi Arabia; however, there is no information about the prevalence of subgenotypes of HBV/D and their relation to advanced liver disease.[17,18] Very few studies related HBV/D1 with severity of disease; however, these studies were hampered by a small sample size and unavailability of balanced clinical groups in comparison.[17,19,20]

Our study was conducted on a cohort of HBV-infected patients in Saudi Arabia recruited from different hospitals in the Kingdom of Saudi Arabia (KSA). We examined HBV genotypes in these patients and analyzed the sequence variations in the EnhII/BCP/precore regions of HBV/D1 associated with the clinical course of the disease. Data on other viral factors, including viral load, HBeAg and antibodies against HBeAg (anti-HBe), were also included to analyze their associations with sequence variations in HBV disease sequelae.

Material and Methods

Patients

A total of 182 serum/plasma samples were obtained from chronic carriers of HBV [presence of HBsAg for >6 months and detection of antibody to hepatitis B core antigen (anti-HBc)] enrolled in different hospitals of KSA. The diagnosis of HCC was based on published guidelines for the diagnosis and management of HCC.[21,22] In brief, enhancement of a liver lesion during the arterial phase and contrast washout during the portal phase in patients with background cirrhosis was considered diagnostic of HCC. Computed tomography and magnetic resonance imaging were the imaging modalities used for diagnosis. Trucut biopsy or fine-needle aspiration was obtained only where considerable doubt existed after imaging studies. Chronic carriers of HBV infection, who had been regularly screened with imaging studies (showing no concerning lesion), performed 6 months apart for 12 months, along with normal α-fetoprotein (AFP) levels were selected as controls. The exclusion criteria for all patients were as follows: (*i*) coinfection with hepatitis C, human immunodeficiency virus or delta virus; (*ii*) coexistent autoimmune or metabolic liver disease; (*iii*) hepatotoxic medications in the preceding 3 months; (*iv*) another hepatobiliary malignancy; (*v*) alcohol consumption >20 g/day and (*vi*) organ transplantation.

Serological markers of infection

Serum samples collected at each hospital were tested for alanine aminotransferase and aspartate aminotransferase (AST), albumin levels and serology for HBeAg and anti-HBe using commercial kits (Abbott laboratories, Diagnostics Division, Abbott Park, IL 60064, USA). HBV genotypes were determined by enzyme-linked immunosorbent assay (ELISA) with monoclonal antibodies directed to distinct epitopes on the preS2-region (HBV GENOTYPE EIA; Institute of Immunology, Tokyo, Japan).

Extraction and quantification of HBV DNA

Total DNA was extracted from 200 μl of serum using a QIAmpDNA mini kit (Qiagen). Quantitative HBV DNA levels were measured by Abbott Real-Time HBV assay (Abbott

Molecular, Des Plaines, IL), allowing detection up to 100 viral DNA copies per milliliter used for DNA quantification.[23]

Amplification of HBV DNA, sequencing and molecular evolutionary analysis

HBV DNA sequences bearing the partial S- and *X*-gene were obtained according to the method proposed by Sugauchi *et al.*[24] with slight modifications. The amplification in the BCP and precore region was carried out with a forward primer HB7F: 5′-GAGACCACCGTGAACGCCCA-3′ (nt. 1611–1630) and an antisense primer HB7R: 5′-CCTGAGTG CAGTATGGTGAGG-3′ (nt. 2072–2052). HBV DNA sequences spanning the S-gene were amplified by two PCR reactions with heminested primers. The first round of PCR was performed with a sense primer HB1F: 5′-AAACTCTGCAAGAT CCCAGAGT-3′ (nt. 18–39) and an antisense primer HB2R: 5′-CAGACTTTCCAATCAATAGG-3′ (nt. 989–970). In the second round, PCR products were obtained in two overlapping fragments. For fragment 1, PCR was performed with the sense primer HB1F and an antisense primer HB1R: 5′-GATACATAGAGGTTCCTTGAGCAG-3′ (nt. 557–534), and for fragment 2, PCR was performed with the sense primer HB2F: 5′-TGCTGCTATGCCTCATCTTC-3′ (nt. 414–433) and the antisense primer HB2R. The amplicons obtained were directly sequenced with Prism Big Dye (Applied Biosystems, Foster City, CA) in an ABI 3100 DNA automated sequencer. HBV genotypes were determined by phylogenetic analysis in the partial S and core regions of HBV genome. Reference sequences for phylogenetic analysis were retrieved from DDBJ/EMBL/GenBank. Alignments were performed using CLUSTALW (http://clustalw.ddbj.nig.ac.jp/top-e.html), and neighbor-joining trees were constructed with six-parametric method and bootstrapped 1,000 times to confirm the reliability of the phylogenetic tree.[25]

Analysis in the EnhII/BCP and precore regions

Analysis of the whole set of HBV/D1-infected cohort was performed in the partial BCP, precore and core regions using BioEdit program version 7.0. The nucleotide mutation was defined by comparing the sequences with differences in the reference HBV/D1 sequence (GenBank accession number AY721612), whereas dual type or deletion was considered as a mutant type. This reference sequence was chosen to be very close to represent the consensus sequence for HBV/D1 based on previously published sequences.

A case–control analysis

Age-matched control patients ($n = 40$) were selected from within the non-HCC cohort with a cutoff age of 60 years. They were matched to within 5 years of the age of case HCC patient. HBeAg was a nonsignificant factor between both groups (p = NS; Table 4). Absence of HCC in the control patient was ascertained by a lack of any definite markers of HCC (normal AFP levels and imaging results showing absence of any concerning lesion) at the point of diagnosis of HCC in the case patient.

Statistical analysis

The nonparametric Mann-Whitney *U* test, Fisher's exact test or χ^2 test with Yate's correction were used to compare data, as appropriate. Stepwise logistic regression analysis was conducted to identify factors independently associated with the development of HCC. All tests were two-sided, and a p-value of less than 0.05 was considered significant. SPSS (version 19) was used to perform the analysis.

Results

Baseline characteristics of patients

The demographic and clinical characteristics of the 182 patients with chronic liver disease (non-HCC = 127 and HCC = 55) are shown in Table 1. Overall, the mean age of the patients was 52.6 ± 20.1 years, and HCC patients were significantly older ($p = 0.0001$) than non-HCC patients. Male gender was significantly abundant among HCC patients ($p = 0.03$). AST, albumin and anti-HBe were significantly higher in HCC patients compared to the non-HCC patients ($p = 0.0001$); inversely, HBeAg positivity was significantly more frequent among non-HCC patients ($p = 0.0001$).

HBV/D was the most prevalent genotype found in 94% (171/182) of patients followed by genotype E in 3.8% (7/182), genotype A2 in 1.6% (3/182) and C2 in 0.5% (1/182) with no statistical significance between both groups. Phylogenetic analysis was used as a major tool to determine the HBV genotypes and subgenotypes. HBV genotyping by ELISA was performed in a total of 18 cases who were either difficult to amplify by PCR or had a short S or core region sequence. These 18 cases along with one subgenotype D2 case were excluded from further analysis. Table 2 presents the baseline characteristics of 152 patients with HBV/D1. Consistent with the findings in the overall cohort (Table 1), age, male gender, anti-HBe, albumin and AST were significantly higher in HCC patients compared to non-HCC patients with HBV/D1 (Table 2).

Patterns of EnhII/BCP and precore mutations

The patterns of the BCP and precore mutations in patients infected with HBV/D1, with or without HCC, are presented in Table 2. A novel double mutation T1673/G1679 located in between Box ά and Box β[10] was found to be significantly higher in the HCC group ($p = 0.007$) compared to the non-HCC group. The frequency of G1727 and C1741 was also significantly higher in the HCC group than in those without HCC ($p = 0.005$ and $p = 0.0006$, respectively). The presence of C1761 was also more frequent in the HCC group ($p = 0.0005$). Apart from this, different kinds of single, double and triple mutation patterns were observed in the region encompassing nucleotides 1757–1768. The polymorphism at position A or G1757 shaped these patterns as double mutation T1762/A1764 and was found in both patterns but did

Table 1. Baseline and clinical characteristics of 182 patients with chronic liver disease infected with HBV in Saudi Arabia

Features	Total ($n = 182$)	Non-HCC ($n = 127$)	HCC ($n = 55$)	p^1
Age[2]	59 (6–93)	56 (6–85)	68 (40–93)	0.0001
Gender (M/F)	131/51	85/42	46/9	0.03
HBeAg+	94 (51.6)	85 (66.9)	9 (16.3)	0.0001
Anti-HBe+	99 (54.3)	50 (39.4)	49 (89)	0.0001
ALT[3]	84.1 ± 154.5	94.6 ± 175	89.5 ± 93.3	NS
AST[3]	101.5 ± 233.3	53.7 ± 144	198 ± 313	0.0001
Albumin[3]	82.1 ± 109.5	42 ± 8.2	183.2 ± 154.9	0.0001
Genotypes				
D	171 (93.9)	119 (93.7)	52 (94.5)	NS
E	7 (3.8)	4 (3.1)	3 (5.4)	NS
A	3 (1.6)	3 (2.3)	0	NS
C	1 (0.5)	1 (0.7)	0	NS

Numbers in parenthesis represent % age.
[1]p: Mann-Whitney U test for continuous data, and χ^2 or Fisher's exact test for categorical data.
[2]Median (range).
[3]Mean ± SD. Abbreviation: ALT: alanine aminotransferase; AST: aspartate aminotransferase; NS: nonsignificant.

not show statistical significance between both clinical groups. However, it was observed that a combination of A1757/C or T1764/G1766 (triple mutation) was significantly higher in HCC patients compared to non-HCC patients ($p = 0.0004$). The point mutation at nucleotide position 1773 (C1773T), alone or in combination with A1775G was significantly higher in HCC patients compared to non-HCC patients ($p = 0.005$ and $p = 0.003$, respectively). The mutations A1896 and A1899 in the precore region were frequent in both groups showing no statistical significance, whereas a novel mutation in the core gene C1909 appeared significantly higher in HCC patients compared to the non-HCC group ($p = 0.014$). Stepwise logistic regression analysis in HBV/D1-infected patients showed older age (>58 years), male gender and viral mutations G1727, T1773, A1757/T1764/G1766 as independent predictive markers of HCC (Table 3).

Age-matched case–control analysis

Age- and gender-matched case–control analysis of 40 patients in each non-HCC and HCC group is shown in Table 4. Cases and controls did not differ significantly for eAg or eAb status in the matched set of samples. There were five controls and two HCC cases that were dually positive for HBeAg and anti-HBe, whereas four controls did not seroconvert. HBV viral load was significantly higher in HCC compared to non-HCC patients. The double mutation T1673/G1679 and the point mutations G1727 and C1741 remained significantly higher in HCC patients ($p = 0.01$, $p = 0.0007$ and $p = 0.006$, respectively). G1757 alone was significantly higher in the non-HCC group ($p = 0.03$), whereas in combination with T1762/A1764, it turned up as a protective mutation pattern, relatively higher in non-HCC compared to HCC patients ($p = 0.08$). The presence of triple mutation A1757/T1764/G1766 in association with HCC was consistent with the overall findings, as shown in Table 1 ($p = 0.01$). As expected, T1773 was more frequent in HCC ($p = 0.0001$), whereas in combination with G1775, it appeared only in the HCC group ($p = 0.01$). Stop-codon mutation A1896 did not reach statistical significance; however, A1899 was associated with non-HCC ($p = 0.04$). The stepwise logistic regression analysis confirmed A1727, A1757/T1764/G1766 and T1773 as independent predictive markers for HCC in this case–control analysis (Table 5).

Discussion

It is believed that HBV genotypes and even subgenotypes may differ in the clinical presentation of the disease and its treatment outcome. In our study, we found that the majority of patients were infected with HBV/D, subgenotype D1 (HBV/D1), followed by genotype E. These results are consistent with previous findings from the region.[13,17] HBV/D1 association with a benign course of disease has been suggested from studies elsewhere; however, most of these studies were hampered by their small sample size. Considering the data arising only from the Mediterranean region, HBV/D7 strains have been reported from Morocco and Tunisia but were not associated with advanced liver diseases.[13,26]

To our knowledge, this is the first case–control study nested within a cohort study of HBV carriers infected with HBV/D1. A significant positive correlation of HBV variants in the *X*/precore region with HCC was found individually and in combination. In the cohort analysis between 107 non-HCC and 45 HCC patients, individual mutations G1727, C1741, C1761 and T1773 were significantly associated with HCC. These mutations, except for C1761, were also associated with HCC in the case–control analysis. The magnitude

Table 2. Comparison of demographic characteristics and BCP and PC mutation factors among 152 patients with chronic liver disease infected with HBV/D1

Factors	Non-HCC ($n = 107$)	HCC ($n = 45$)	p^1
Age[2]	56 (26–65)	68 (65.5–71.5)	0.0001
Gender (M/F)	71/36	39/6	0.01
HBeAg+	66 (61.6)	5 (11.1)	0.0001
Anti-HBe+	47 (43.9)	41 (91.1)	0.0001
ALT[3]	86.5 ± 187.8	93.5 ± 94.5	NS
AST[3]	69.7 ± 197.9	206.9 ± 333.7	0.002
AFP[3]	3 (2.6–7.3)	36.5 (6.8–1,000)	0.0001
Albumin[3]	40.9 ± 8.02	186.6 ± 165.4	0.0001
Viral load (log IU/ml)[3]	6.03 ± 3.3	6.65 ± 3.7	NS
C1653 T or Y	14 (13)	8 (17.7)	NS
T1678 C	10 (9.3)	4 (8.8)	NS
A1679 G (alone)	10	3	0.08
C1673T/A1679G (double)	3 (2.8)	7 (15.5)	0.007
A1727 G	13 (12.1)	15 (33.3)	0.005
T1741 C	6 (5.6)	12 (26.6)	0.0006
T1753 C or A	43 (40.1)	16 (35.5)	NS
A1757G	38 (35.5)	12 (26.6)	NS
T1761C	2 (1.8)	8 (17.7)	0.0005
A1762T (alone)	5 (4.6)	1 (2.2)	NS
G1764A or T (alone)	4 (3.7)	0	ND
G1757/T1762/A1764 (triple)	26 (24.2)	9 (20)	NS
A1757/T1762/A1764 (triple)	13 (12.1)	4 (8.8)	NS
A1757/T or C1764/G1766 (triple)	12 (11.2)	16 (35.6)	0.0004
C1766T (alone)	7 (6.5)	5 (11.1)	NS
C1766T/T1768A (double)	9 (8.4)	5 (11.1)	NS
C1773T (alone)	60 (56.0)	36 (80)	0.005
A1775G (alone)	4 (3.7)	0	ND
C1773 T/A1775G (double)	2 (1.8)	7 (15.5)	0.003
G1896A	47 (43.9)	25 (55.5)	NS
G1899A	37 (34.5)	19 (42.2)	NS
T1909C	8 (7.4)	10 (22.2)	0.014
T1912C	10 (9.3)	2 (4.4)	NS

Numbers in parenthesis represent % age.
[1]p: Mann-Whitney U test for continuous data, and χ^2 or Fisher's exact test for categorical data.
[2]Median (interquartile range).
[3]Mean ± SD. Abbreviations: ALT: alanine aminotransferase; AST: aspartate aminotransferase; NS: nonsignificant; ND: not determined.

of OR was highest for T1773 mutation (silent mutation) in the cohort and the case–control analysis, that is, 11.8 and 14, respectively. The presence of this mutation in severe liver disease has also recently been reported among Turkish patients infected with HBV/D1.[27] The clinical impact of the point mutation A1727G (silent mutation) is not clear in HBV/D1 infection; however, the reverse mutation, G1727A, has been reported as a marker of HCC in occult HBV infection from Taiwan where infections by genotypes B and C are common.[12] The missense mutation C1761, causing amino acid change K130Q, has previously been reported from Iran in connection with severe liver disease.[19] The missense point mutation C1741 causing amino acid change L123S is novel; however, the mechanism whereby its interaction exists with other BCP mutations is yet unclear.

Our observation of an increased risk of HCC associated with infection by HBV strains in combination of mutations in the X-gene is far more novel and interesting. In the case–control analysis, the missense point mutation C or A1753 (causing amino acid change I1127N/T) appears to be a

protective one ($p = 0.03$), in contrast to the previous finding where this mutation has been found associated with the development of HCC in HBV/C-infected patients. Polymorphism at nucleotide position 1757 (sense mutation) has been evidenced in relation with the BCP double mutations T1762/A1764 (CP1) or T1764/G1766 (CP2).[19,28,29] In our study, the CP1 mutation appeared in combination with G at position 1757, showing a protective trend from HCC ($p = 0.08$); however, this needs to be further studied and confirmed by larger studies. CP1 affects amino acid changes K130M and V131I and, contrary to our findings, is a characteristic HCC-related double mutation in HBV/C and/or HBV/Ba infections.[6,10,30] The CP2 mutation that results in amino acid change C131L was observed in combination with A1757 and was significantly associated with HCC ($p = 0.01$). Our results are in agreement with previous reports where this double mutation has been reported in association with severe liver disease in HBV/D infections.[19,28] A recent *in vitro* study[29] showed that the CP2 mutation induced high levels of viral replication and transcription efficiency in HuH7 and HepG2 cells, which were comparable to those induced by the CP1 mutation. The effect of the CP2 mutation was significantly increased by the addition of the 1757A mutation by creating a binding site for the transcription factor HNF3, thereby increasing its

Table 3. Stepwise logistic regression analysis for factors independently associated with the development of HCC in patients infected with HBV/D1

Factors	Odds ratio (95% CI)	p^1
Age (>58 years)	6.78 (2.5–18.3)	0.0001
Male	2.96 (0.93–9.4)	0.06
G1727	3.97 (1.34–11.7)	0.01
T1764/G1766	2.8 (1.01–7.8)	0.04
T1773	11.8 (2.5–55.7)	0.002

[1]p: Wald test.

Table 4. Age- and gender-matched case–control analysis in the BCP and PC regions of HBV/D1 in patients with chronic liver disease

Factors	Non-HCC ($n = 40$)	HCC ($n = 40$)	p^1
HBeAg	4 (10)	0	NS
HBeAg + anti-HBe	5 (12.5)	2 (5)	NS
Viral load (log_{IU} ml^{-1})[2]	4.33 ± 3.0	6.48 ± 3.9	0.008
C1653T	6 (15)	7 (17.5)	NS
A1679G (alone)	3 (7.5)	3 (7.5)	0.06
C1673T/A1679G (double)	0	7 (17.5)	0.01
T1678C	8 (20)	4 (10)	NS
A1727G	2 (5)	14 (35)	0.0007
T1741C	3 (7.5)	12 (30)	0.006
T1753C or A	21 (52.5)	12 (30)	0.07
A1757G	21 (52.5)	11 (27.5)	0.03
A1761C	1 (2.5)	6 (15)	NS
A1762T (alone)	3 (7.5)	1 (2.5)	NS
G1764A or T (alone)	1 (2.5)	1 (2.5)	NS
G1757G/T1762T/A1764 (triple)	16 (40)	8 (20)	0.08
A1757/T1762/A1764 (triple)	4 (10)	3 (7.5)	NS
A1757/T or C1764/G1766 (triple)	5 (12.5)	15 (37.5)	0.01
C1766T (alone)	2 (5)	3 (7.5)	NS
C1766T/T1768A (double)	3 (7.5)	3 (7.5)	NS
C1773T	21 (52.5)	38 (95)	0.0001
C1773T/A1775G (double)	0	7 (17.5)	0.01
G1896A	29 (72.5)	22 (55)	NS
G1899A	24 (60)	16 (40)	0.04
T1909C	5 (12.5)	9(20)	NS
T1912C	4 (12.5)	2 (5)	NS

Numbers in parenthesis represent % age.
[1]p: Mann-Whitney U test for continuous data, and χ^2 or Fisher's exact test for categorical data.
[2]Mean ± SD. Abbreviation: NS: nonsignificant.

Table 5. Stepwise logistic regression analysis for factors independently associated with HCC development for age- and gender-matched case–control subjects

Factors	Odds ratio (95% CI)	p^1
G1727	18.3 (2.8–118.4)	0.002
T1764/G1766	4.7 (1.31–17.2)	0.01
T1773	14.06 (2.3–84.8)	0.004

[1]p: Wald test.

transcriptional activity. In contrast, introduction of the 1757A mutation reduced the transcriptional activity of CP1, abolishing the viral replication *via* a reduction in HNF1 binding affinity. The double mutation T1766/A1768 was found in a small population of HCC patients and controls. The exact significance of this double mutation is not entirely clear; however, a few studies have described it as a predictive marker for cirrhosis.[31] Apart from these double mutations, a quadruple mutation T1673/G1679/T1773/G1775 was observed in a group of seven HCC patients. Being silent mutations, the exact biological significance of this combination of mutations is not entirely clear. Interestingly, these cases were also carrying point mutations G1727 and C1741. It is possible that G1727 and C1741 interact with the quadruple mutation in a similar way as CP1 and CP2 mutations interact with G or A1757, affecting the transcription factor binding site and inducing high levels of viral replication. It is possible that silent and missense mutations may synergistically act for a significantly altered function of X-protein, promoting hepatocarcinogenesis by interfering with cell growth control and DNA repair. According to a previous study, there may be a dose-risk relationship of mutation number with HCC and suggested using the mutation count as a diagnostic indicator for HCC.[14]

The G to A change at position 1896 is a hot-spot mutation in the precore region, which creates a premature stop-codon and has been associated with HBeAg levels.[32] Inconsistent results have been reported about the relationship of this mutation with liver disease. It has been associated with fulminant hepatitis in some studies[33,34] or less hepatic inflammation,[33] whereas some other studies did not find any notable association with liver disease.[35–38] Our study findings showed its association with HBeAg seroconversion, but could not relate it to the development of HCC. Furthermore, the accumulating evidence suggests that HBV/D exists more as HBeAg-negative phenotype. Various patients do seroconvert in the initial stages of infection, although not clearing the virus itself but remaining a carrier for life, suggesting an immune selection phenomenon as opposed to a replication advantage.[38,39]

In conclusion, we have shown several novel mutations in the EnhII/BCP regions of the HBV genome associated with the development of HCC. Each specific mutation may be sufficiently associated with HCC; however, the synergistic effect of combination patterns of mutations may be much more critical in escalating the development of HCC. These mutation complexes are novel risk factors that may facilitate early prediction of HCC in the chronic carriers of HBV/D1 infection.

References

1. Liaw YF, Chu CM. Hepatitis B virus infection. *Lancet* 2009;373:582–92.
2. Bruix J, Sherman M. Management of hepatocellular carcinoma. *Hepatology* 2005;42:1208–36.
3. Orito E, Ichida T, Sakugawa H, et al. Geographic distribution of hepatitis B virus (HBV) genotype in patients with chronic HBV infection in Japan. *Hepatology* 2001;34:590–4.
4. Tatematsu K, Tanaka Y, Kurbanov F, et al. A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. *J Virol* 2009;83:10538–47.
5. Yu H, Yuan Q, Ge SX, et al. Molecular and phylogenetic analyses suggest an additional hepatitis B virus genotype "I". *PLoS One* 2010;5:e9297.
6. Chan HL, Hui AY, Wong ML, et al. Genotype C hepatitis B virus infection is associated with an increased risk of hepatocellular carcinoma. *Gut* 2004;53:1494–8.
7. Ito K, Tanaka Y, Orito E, et al. T1653 mutation in the Box-α increases the risk of hepatocellular carcinoma in patients with chronic hepatitis B virus genotype C infection. *Clin Infect Dis* 2006;42:1–7.
8. Sakamoto T, Tanaka Y, Simonetti J, et al. Classification of hepatitis B virus genotype B into 2 major types based on characterization of a novel subgenotype in Arctic indigenous populations. *J Infect Dis* 2007;196:1487–92.
9. Lau JY, Wright TL. Molecular virology and pathogenesis of hepatitis B. *Lancet* 1993;342:1335–40.
10. Tanaka Y, Mukaide M, Orito E, et al. Specific mutations in enhancer II/core promoter of hepatitis B virus subgenotypes C1/C2 increase the risk of hepatocellular carcinoma. *J Hepatol* 2006;45:646–53.
11. Chou YC, Yu MW, Wu CF, et al. Temporal relationship between hepatitis B virus enhancer II/basal core promoter sequence variation and risk of hepatocellular carcinoma. *Gut* 2008;57:91–7.
12. Chen CH, Changchien CS, Lee CM, et al. A study on sequence variations in pre-S/surface, X and enhancer II/core promoter/precore regions of occult hepatitis B virus in non-B, non-C hepatocellular carcinoma patients in Taiwan. *Int J Cancer* 2009;125:621–9.
13. Kitab B, Essaid El Feydi A, Afifi R, et al. Variability in the precore and core promoter regions of HBV strains in morocco: characterization and impact on liver disease progression. *PLoS One* 2012;7:e42891.
14. Jang JW, Chun JY, Park YM, et al. Mutational complex genotype of the hepatitis B virus X/precore regions as a novel predictive marker for hepatocellular carcinoma. *Cancer Sci* 2012;103:296–304.
15. Chandra PK, Biswas A, Datta S, et al. Subgenotypes of hepatitis B virus genotype D (D1, D2, D3 and D5) in India: differential pattern of mutations, liver injury and occult HBV infection. *J Viral Hepat* 2009;16:749–56.
16. Yousif M, Kramvis A. Genotype D of hepatitis B virus and its subgenotypes: an update. *Hepatol Res* 2013;43:355–64.
17. Abdo AA, Al-Jarallah BM, Sanai FM, et al. Hepatitis B genotypes: relation to clinical outcome in patients with chronic hepatitis B in Saudi Arabia. *World J Gastroenterol* 2006;12:7019–24.
18. Sanai FM, Helmy A, Bzeizi KI, et al. Discriminant value of serum HBV DNA levels as predictors of liver fibrosis in chronic hepatitis B. *J Viral Hepatitis* 2011;18:e217–e225.
19. Veazjalali M, Norder H, Magnius L, et al. A new core promoter mutation and premature stop codon in the S gene in HBV strains from Iranian patients with cirrhosis. *J Viral Hepat* 2009;16:259–64.
20. Sendi H, Mehrab-Mohseni M, Zali MR, et al. T1764G1766 core promoter double mutants are restricted to Hepatitis B virus strains with an A1757 and are common in genotype D. *J Gen Virol* 2005;86:2451–8.
21. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology* 2011;53:1020–2.
22. Abdo AA, Hassanain M, Al Jumah A, et al. Saudi guidelines for the diagnosis and management of hepatocellular carcinoma: technical review and practice guidelines. *Ann Saudi Med* 2012;32:174–99.

23. Tanaka Y, Hasegawa I, Kato T, et al. A case–control study for differences among hepatitis B virus infections of genotypes A (subtypes Aa and Ae) and D. *Hepatology* 2004;40:747–55.
24. Sugauchi F, Mizokami M, Orito E, et al. A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. *J Gen Virol* 2001;82:883–92.
25. Shin IT, Tanaka Y, Tateno Y, et al. Development and public release of a comprehensive hepatitis virus database. *Hepatol Res* 2008;38:234–43.
26. Hannachi N, Fredj NB, Bahri O, et al. Molecular analysis of HBV genotypes and subgenotypes in the Central-East region of Tunisia. *Virol J* 2010;7:302.
27. Sunbul M, Sugiyama M, Kurbanov F, et al. Specific mutations of basal core promoter are associated with chronic liver disease in hepatitis B virus subgenotype D1 prevalent in Turkey. *Microbiol Immunol* 2013;57:122–9.
28. Elkady A, Tanaka Y, Kurbanov F, et al. Virological and clinical implication of core promoter C1752/V1753 and T1764/G1766 mutations in hepatitis B virus genotype D infection in Mongolia. *J Gastroenterol Hepatol* 2008;23:474–81.
29. Sugiyama M, Tanaka Y, Sugauchi F, et al. Hepatitis virus genotype D1 specific mutations affecting viral replication and clinical course of hepatitis B patients. HBV Now in Asia: Single Topic Conference. 2009. 85.
30. Yu MW, Yeh SH, Chen PJ, et al. Hepatitis B virus genotype and DNA level and hepatocellular carcinoma: a prospective study in men. *J Natl Cancer Inst* 2005;97:265–72.
31. Guo X, Jin Y, Qian G, et al. Sequential accumulation of the mutations in core promoter of hepatitis B virus is associated with the development of hepatocellular carcinoma in Qidong, China. *J Hepatol* 2008;49:718–25.
32. Tacke F, Gehrke C, Luedde T, et al. Basal core promoter and precore mutations in the hepatitis B virus genome enhance replication efficacy of Lamivudine-resistant mutants. *J Virol* 2004;78:8524–35.
33. Liang TJ, Hasegawa K, Rimon N, et al. A hepatitis B virus mutant associated with an epidemic of fulminant hepatitis. *N Engl J Med* 1991;324:1705–9.
34. Hasegawa K, Huang JK, Wands JR, et al. Association of hepatitis B viral precore mutations with fulminant hepatitis B in Japan. *Virology* 1991;185:460–3.
35. Kao JH, Chen PJ, Lai MY, et al. Basal core promoter mutations of hepatitis B virus increase the risk of hepatocellular carcinoma in hepatitis B carriers. *Gastroenterology* 2003;124:327–34.
36. Fang ZL, Yang J, Ge X, et al. Core promoter mutations (A(1762)T and G(1764)A) and viral genotype in chronic hepatitis B and hepatocellular carcinoma in Guangxi, China. *J Med Virol* 2002;68:33–40.
37. Chu CJ, Keeffe EB, Han SH, et al. Prevalence of HBV precore/core promoter variants in the United States. *Hepatology* 2003;38:619–28.
38. Fang ZL, Sabin CA, Dong BQ, et al. The association of HBV core promoter double mutations (A1762T and G1764A) with viral load differs between HBeAg positive and anti-HBe positive individuals: a longitudinal analysis. *J Hepatol* 2009;50:273–80.
39. Livingston SE, Simonetti JP, Bulkow LR, et al. Clearance of hepatitis B e antigen in patients with chronic hepatitis B and genotypes A, B, C, D, and F. *Gastroenterology* 2007;133:1452–7.

JOURNAL OF VIRAL HEPATITIS

Journal of Viral Hepatitis, 2013, **20**, e27–e36 doi:10.1111/jvh.12022

Mechanism of the dependence of hepatitis B virus genotype G on co-infection with other genotypes for viral replication

T. Sakamoto,[1,2] Y. Tanaka,[1] T. Watanabe,[1] S. Iijima,[1] S. Kani,[1] M. Sugiyama,[1] S. Murakami,[1] K. Matsuura,[1,2] A. Kusakabe,[1,2] N. Shinkai,[1,2] F. Sugauchi[1,2] and M. Mizokami[1,3] [1]*Department of Virology & Liver unit;* [2]*Department of Gastroenterology and Metabolism, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; and* [3]*Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan*

Received April 2012; accepted for publication September 2012

SUMMARY. Hepatitis B virus (HBV) is classified into several genotypes. Genotype G (HBV/G) is characterised by worldwide dispersion, low intragenotypic diversity and a peculiar sequence of the precore and core region (stop codon and 36-nucleotide insertion). As a rule, HBV/G is detected in co-infection with another genotype, most frequently HBV/A2. In a previous *in vivo* study, viral replication of HBV/G was significantly enhanced by co-infection with HBV/A2. However, the mechanism by which co-infection with HBV/A2 enhances HBV/G replication is not fully understood. In this study, we employed 1.24-fold HBV/A2 clones that selectively expressed each viral protein and revealed that the core protein expressing construct significantly enhanced the replication of HBV/G in Huh7 cells. The introduction of the HBV/A2 core promoter or core protein or both genomic regions into the HBV/G genome showed that both the core promoter and core protein are required for efficient HBV/G replication. The effect of genotype on the interaction between foreign core protein and HBV/G showed that HBV/A2 was the strongest enhancer of HBV/G replication. Furthermore, Western blot analysis of Dane particles isolated from cultures of Huh7 cells co-transfected by HBV/G and a cytomegalovirus (CMV) promoter–driven HBV/A2 core protein expression construct indicated that HBV/G employed HBV/A2 core protein during particle assembly. In conclusion, HBV/G could take advantage of core proteins from other genotypes during co-infection to replicate efficiently and to effectively package HBV DNA into virions.

Keywords: co-transfection, core protein, genotype A, genotype G, hepatitis B virus, replication.

INTRODUCTION

Hepatitis B virus (HBV) infection affects more than 350 million people and is one of the major causes of acute and chronic liver disease. Acute HBV infection in adults is usually self-limiting, while chronic HBV infection can cause chronic hepatitis, liver cirrhosis or hepatocellular carcinoma [1]. As the clinical course in infected individuals depends on a complex interplay among various factors including viral, host and environmental factors, molecular characteristics of HBV including the genotype could become increasingly important in our understanding of HBV clinical implications [2].

Abbreviations: CMV, cytomegalovirus; CP, core promoter; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; SEAP, secreted alkaline phosphatase.

Correspondence: Masashi Mizokami, MD, PhD, The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine 1-7-1, Kohnodai, Ichikawa 272-8516, Japan. E-mail: mmizokami@hospk.ncgm.go.jp

Eight major HBV genotypes (A–H) have been identified by a sequence divergence >8% in the entire HBV genome [3,4] and have a relatively distinct geographical distribution, which may be associated with anthropological history [5]. Hepatitis B virus genotype G (HBV/G) was first described in 2000 by studies carried out in France [6]. It is usually detected during co-infection with other genotypes, most frequently with HBV/A2 [7,8]. Co-infection with HBV/C and H has also been reported [9–11]. One of the features distinguishing HBV/G from other genotypes is the 36-nucleotide (nt) insertion in its core gene [6,12]. Recent studies indicated that the 36-nt insertion increased core protein translation without enhancing mRNA abundance [13], and insertion of the 36-nt in the core region of genotypes A and D impaired genome replication, despite upregulation of core protein expression, indicating that the 36-nt insertion could alter core protein expression without altering the mRNA expression [14]. The other feature of the HBV/G genome that is unique is the possession of two stop codons in the precore region that prevents the expression of hepatitis B e antigen (HBeAg) [6,12]. Nevertheless, some HBV/G carriers are