Mol. Cells *29*, 123-130, February 28, 2010
DOI/10.1007/s10059-010-0008-0

Molecules and Cells

Mining of Serum Glycoproteins by an Indirect Approach Using Cell Line Secretome

Younghee Ahn[1,3], Un-Beom Kang[1,2,3], Joon Kim[2], and Cheolju Lee[1,*]

Glycosylation is the most important and abundant post-translational modification in serum proteome. Several specific types of glycan epitopes have been shown to be associated with various types of disease. Direct analysis of serum glycoproteins is challenging due to its wide dynamic range. Alternatively, glycoproteins can be discovered in the secretome of model cell lines and then confirmed in blood. However, there has been little experi-mental evidence showing cell line secretome as a tractable target for the study of serum glycoproteins. We used a hydrazine-based glycocapture method to selectively enrich glycoproteins from the secretome of the breast cancer cell line Hs578T. A total of 132 glycoproteins were identified by nanoLC-MS/MS analysis. Among the identified proteins, we selected 13 proteins that had one or more *N*-glycosylation motifs in the matched peptides, which were included in the Secreted Protein Database but not yet in the Plasma Proteome Database (PPD), and whose antibodies were commercially available. Nine out of the 13 selected proteins were detected from human blood plasma by western analysis. Furthermore, eight proteins were also detected from the plasma by targeted LC-MS/MS, which had never been previously identified by data-dependent LC-MS/MS. Our results provide novel proteins that should be enrolled in PPD and suggest that analysis of cell line secretome with subfractionation is an efficient strategy for discovering disease-relevant serum proteins.

INTRODUCTION

The dynamic nature of the human circulatory system and its constituents reflects diverse physiological or pathological states, and the ease with which blood can be sampled makes it a logical choice for biomarker applications. Plasma and serum are the most informative proteome source from a medical viewpoint since many cells cross communicate through the secretion of soluble proteins into blood (Amin et al., 2006; Siriwardana et al., 2006). Proteins that are detectable in serum and plasma form the basis of commonly used tests to screen and monitor various diseases through the measurement of disease-specific antigen (Hanash et al., 2008). In the past few years, several large-scale proteomic studies of human diseases using plasma have begun to characterize the disease-relevant proteome, and proteomic methods based on mass spectrometry have matured significantly and promise to deliver candidate markers for diagnosis, prognosis, or monitoring therapeutic response. This high-throughput strategy leads to complex data sets, and while rich in information, it is often difficult to use to predict proteins that may be sensitive and specific biomarkers for the disease. The ability to add an additional, complementary means to enrich for potential marker proteins would increase the probability of identifying candidate biomarkers.

Glycosylation is clearly the most complex set of post-translational modifications that proteins undergo during biosynthesis, and several specific types of glycan epitopes have been shown to be associated with various types of cancer (Hakomori, 2001; Kim and Varki, 1997). The enrichment of *N*-glycoproteins is of particular interest for characterizing the plasma proteome because the majority of plasma proteins are believed to be glycosylated. The changes in abundance and the alternations in glycan composition of plasma proteins and cell surface proteins have been shown to correlate with cancer and other disease states. In fact, numerous clinical biomarkers and therapeutic targets are glycosylated proteins, such as the prostate-specific antigen for prostate cancer, and CA125 for ovarian cancer (Roth, 2002). Recently, Zhang et al. (2003) developed an approach for specific enrichment of N-linked glycopeptides using hydrazide chemistry. Because the *N*-glycosylation sites generally fall into a consensus NXS/T sequence motif in which X represents any amino acid residue except proline (Bause, 1983), this motif can be used as a sequence tag prerequisite to aid in confident validation of N-glycopeptide identifications during mass spectrometry (Liu et al., 2005).

Although human serum represents an important biological material for disease diagnosis, the wide, dynamic range in protein concentration imposes a major hurdle in the direct analysis of serum proteome by mass spectrometry. Furthermore, confident identification and quantification for the very low-concentration proteins is a great challenge in the presence of high-abundance proteins. A secretome, composed of secreted proteins from a model cell line, may serve as a reservoir for serological markers if some of the proteins are actually secreted

[1]Life Sciences Division, Korea Institute of Science and Technology, Seoul 136-791, Korea, [2]School of Life Sciences and Biotechnology, Korea University, Seoul 136-701, Korea, [3]These authors contributed equally to this work.
*Correspondence: clee270@kist.re.kr

Received July 20, 2009; revised October 12, 2009; accepted October 27, 2009; published online December 10, 2009

Keywords: glycoprotein, LC-MS/MS, plasma proteome database, proteome, secreted protein database, secretome

from the disease site and smeared into the circulation. In such a case, analysis of cell line secretome followed by validation in human plasma or serum may be an alternative way for biomarker discovery. In spite of several successful examples of this indirect approach (Chang et al., 2008; Huang et al., 2006; Wu et al., 2005), there was no systematic study by experiment to prove the supposition that a secretome from a disease-model cell line is a reservoir of serum proteins.

In this study, we enriched and analyzed *N*-glycoproteins from the secretome of Hs578T breast cancer cell line by using 2D LC-MS/MS. A total of 132 proteins were found in the conditioned media, among which eight proteins were confirmed for their existence in blood plasma by western analysis and targeted LC-MS/MS. Our result suggests that cell line secretome is a useful source for the discovery of low abundant serum proteins. The identified proteins may also have important implications for the eventual development of clinical diagnostic tests using plasma and for our understanding of cancer biology.

MATERIALS AND METHODS

Cell culture

The Hs578T cell line was obtained from the American Type Culture Collection. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen, USA), supplemented with 10% fetal bovine serum (Invitrogen), 10 μg/ml of insulin (Sigma, USA), 100 unit/ml of penicillin (Invitrogen), and 100 μg/ml of streptomycin (Invitrogen), at 37°C in a humidified 95% air, 5% CO_2 incubator. Cells were washed with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 2 mM KH_2PO_4, pH 7.4) five times at the 90% confluence stage, and then incubated for 6 h in a serum-free conditioned medium. Secreted proteins were harvested by collecting the conditioned media and concentrated by ultrafiltration with Centricon (3,000 Da cutoff; Millipore, USA). Protein concentration was determined with a BCA protein assay kit (Pierce, USA).

Enrichment of tryptic N-glycopeptides

Hydrazide resin (Bio-Rad, USA) was used to capture glycoproteins as described (Zhang et al., 2003). The concentrated secretome was diluted 10-fold in a coupling buffer (100 mM sodium acetate and 150 mM NaCl, pH 5.5) and oxidized in 15 mM sodium periodate at 25°C for 1 h in the dark with constant shaking. The sodium periodate was subsequently removed by using a pre-packed PD-10 column (GE Healthcare, USA) equilibrated with the coupling buffer. The hydrazide resin was washed five times with the coupling buffer, and the oxidized protein sample was then added and incubated with the resin overnight at 25°C. Nonglycoproteins were removed by washing the resin briefly three times with 100% methanol and then three times with 8 M urea in 0.4 M NH_4HCO_3. The glycoproteins were denatured in 8 M urea and reduced with 10 mM DTT at 37°C for 1 h. Protein cysteinyl residues were alkylated with 20 mM iodoacetamide for 90 min at 25°C. After washing with 8 M urea and 50 mM NH_4HCO_3, the resin was resuspended as a 20% slurry in 50 mM NH_4HCO_3 and sequencing grade trypsin (Promega, USA) was added at a 1:100 (w:w) trypsin-to-protein ratio. The sample was digested on-resin overnight at 37°C. The trypsin-released peptides were removed by washing the resin extensively with three separate solutions: 2 M NaCl, 100% methanol, and 50 mM NH_4HCO_3. The resin was resuspended as 50% slurry in 50 mM NH_4HCO_3 and *N*-glycopeptides were released by incubating the resin with 0.3 μl of peptide-N-glycosidase F (PNGase F, 500 unit/μl; New England Biolabs, USA) for 4 h at 37°C. The released glycopeptides were cleaned using a SPE C_{18} column (Supelco, USA) and dried under vacuum.

Peptide fractionation by strong cation exchange chromatography

Dried glycopeptides were reconstituted in 200 μl of 10 mM KH_2PO_4 (pH 3.0) in 25% acetonitrile and fractionated by using a Polysulfoethyl-A SCX column (200 × 4.6 mm, 5 μm, 300 Å; PolyLC, USA). The separation was performed at a flow rate of 0.8 ml/min using an ÄKTA Explorer System (Amersham Bioscience, Sweden). After loading the sample onto the column, peptides were then eluted with a linear gradient from 0 to 0.4 M KCl over 40 min. A total of 7 fractions were collected, and each fraction was desalted by using a C_{18} SPE cartridge, and dried under vacuum.

Reversed-phase nano LC-MS/MS analysis

The dried SCX fractions were dissolved in 0.4% acetic acid and analyzed by nanoLC-MS/MS. A binary nanoflow HPLC system (Agilent 1200 series; Agilent, USA) was directly coupled to an LTQ linear ion trap mass spectrometer (Thermo Electron Corp., USA) with automatic gain control to avoid space charge limitations. In-house packed reversed-phase column (5-μm C_{18} material, column dimension 75 μm × 13 cm) was used at a flow rate of 400 nl/min. Gradient elution of the peptide sample was achieved using 90% solvent A (0.1% formic acid in H_2O) to 40% solvent B (0.1% formic acid in acetonitrile) over 85 min. The MS survey was scanned from 300 to 2,000 *m/z*, and followed by three data-dependent MS/MS scans with the followed options: isolation width, 1.5 m/z; normalized collision energy, 25%; dynamic exclusion duration, 180 s.

Database search

For the analysis of tandem mass spectral data, we constructed a data analysis pipeline system with both a SEQUEST database search engine and Trans-Proteomic Pipeline (TPP; Version 2.9; INTERACT, PeptideProphet™, ProteinProphet™) provided from the Institute for Systems Biology (Keller et al., 2005). SEQUEST searches were performed against the human International Protein Index database (IPI, versions 3.13, www.ebi.ac.uk/IPI/). Three trypsin-missed cleavages were allowed and the peptide mass tolerances for MS/MS and MS were set to ± 0.5 and ± 3 Da, respectively. Other options used for SEQUEST searches were fixed modification of carbamidomethylation at cysteine (+ 57.0215 Da), potential modification of oxidation at methionine (+ 15.9949 Da), and increment of 0.9848 Da for each asparagine glycosylation site. Proteins whose ProteinProphet probability was greater than 0.9 were collected.

Depletion of high-abundance plasma proteins

Plasmas collected from six breast cancer patients (Chang et al., 2008) were pooled and diluted (1:5) with a proprietary "Buffer A" from Agilent Technologies (Agilent) as part of their antibody-based depletion system for highly abundant proteins. Six high-abundant plasma proteins, albumin, IgG, antitrypsin, IgA, transferrin, and haptoglobin, which constitute approximately 85% of the total protein mass of human plasma, were removed in a single step by using a MARS affinity column (Agilent) on an Agilent 1100 series HPLC system (Agilent) according to the manufacturer's instructions. The unbound fraction was concentrated using a Microcon ultrafiltration system (3,000 Da cut off, Millipore).

Western analysis

Plasma samples (5 μg protein) were separated by SDS-PAGE, transferred onto a nitrocellulose membrane (Bio-Rad), blocked

with 5% skim milk, and then probed with appropriate antibodies. The primary antibodies used in this study include the following antibodies directed against glutathione peroxidase 3 (Abcam, Cambridge, USA), fetuin-b (R&D system, Minneapolis, USA), FK506-binding protein 10 (Abnova, Taipei, Taiwan), prosaposin (Abnova), bone morphogenetic protein 1 (Abcam), insulin-like growth factor binding protein 3 (Santa Cruz, USA), tyrosine-protein kinase TXK (Abcam), protein-lysine 6-oxidase (Abcam), calumenin (Santa Cruz), cathepsin L (Abcam), and N-acetylglucosamine-6-sulfatase (R&D system). Blots were wa-shed three times with TBST buffer (20 mM Tris, 150 mM NaCl, and 0.05% Tween 20, pH 7.4), probed with horseradish peroxidase-conjugated secondary antibody for 1 h at 25°C, and then developed with a chemiluminescence detection system (GE Healthcare and Pierce).

In-gel digestion

Bands in the gel were excised and destained in 25 mM NH_4HCO_3 and 50% acetonitrile. Proteins in the bands were reduced with 10 mM DTT in 25 mM NH_4HCO_3 at 56°C for 1 h, and alkylated with 55 mM iodoacetamide at 25°C for 1 h in the dark. The gel pieces were washed with 25 mM NH_4HCO_3 and 50% acetonitrile, and soaked in a chilled 25 mM NH_4HCO_3 solution containing 12 ng/μl of sequencing grade trypsin (Promega) for 45 min. After removing residual trypsin, the gel pieces were incubated at 37°C for 16 h. Digested peptides were extracted sequentially with 25 mM NH_4HCO_3 in 50% acetonitrile, 0.25% TFA in 50% acetonitrile, and 0.25% TFA in 70% acetonitrile. The three extracts were pooled and dried by using a vacuum evaporator.

Targeted LC-MS/MS

Targeted LC-MS/MS for predetermined precursors was performed on the LTQ linear ion-trap mass spectrometer. The target precursors were chosen from the Global Proteome Machine (GPM) database (http://gpmdb.thegpm.org/). From the proteotypic peptides representing the target proteins, precursors between 500 and 1,000 *m/z* at the charge state of +1, +2, or +3 were generally selected. Mass spectra were acquired by a procedure of alternating an MS scan over the m/z range 300-2,000 and an MS/MS scan for the predetermined precursor ion. The LC-MS/MS parameters were as follows: isolation width, 2 *m/z*; normalized collision energy, 35% (Lee et al., 2004).

RESULTS AND DISCUSSION

In this study, we focused on the identification of *N*-glycoproteins secreted from a cancer cell line and further confirmation of their existence in human plasma. The background rationale for the choice of *N*-glycosylation is as follows: 1) *N*-glycosylation is particularly prevalent in blood plasma (Roth, 2002), 2) the overall specificity of *N*-glycopeptide capture through hydrazide chemistry and release by PNGase F hydrolysis has been well demonstrated (Zhang et al., 2005), and 3) the selective isolation of the *N*-linked glycosylated peptides from the milieu of tryptic peptides will result in a substantial improvement in the detection limit of protein due to the reduction in sample complexity compared to whole glycoprotein analysis. Secreted proteins are also targeted in many biomarker studies because they are more likely to be present in body fluids such as serum or plasma (Klee et al., 2006). We used secretome of malignant breast cancer cells, Hs578T. With the secretome from the cell line, we succeeded previously in discovering a novel breast cancer biomarker, as well as several low-abundant serum proteins (Chang et al., 2008).

Analysis of N-glycoproteins secreted from the Hs578T cell line

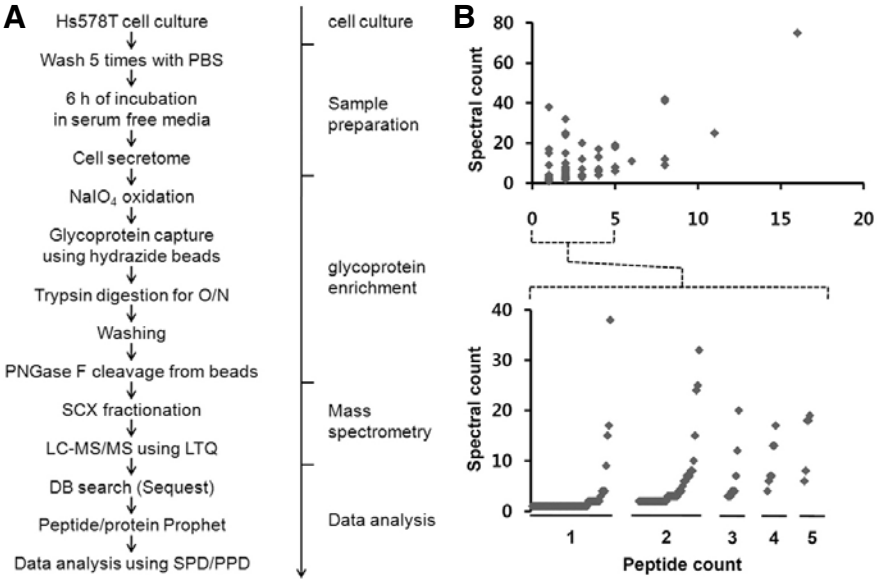
Almost all cancer cell lines must be maintained and cultured in the presence of serum factors. As Hs578T was not an exception to this, the cells were extensively washed in PBS to remove any contaminants originating from serum in the medium. The cells were washed 5 times before being transferred to conditioned media. Less than 5 washes left too much serum protein and more than 5 washes affected the viability of the cells (Chang et al., 2008; Lawlor et al., 2009). A 6-hour incubation in serum free media was used since the concentration of secretome was relatively high without affecting cell viability to a significant degree (Chang et al., 2008).

The entire workflow of this approach is illustrated in Fig. 1A. Two-dimensional chromatography consisting of SCX and C_{18} reversed phase LC was employed to improve the dynamic range of detection effectively and increase the coverage for low-abundance proteins. Glycopeptides released by PNGase F were fractionated into 7 fractions using SCX chromatography, and each fraction was analyzed by LC-MS/MS. Since the PNGase F catalyzed deglycosylation reaction converts each asparagine to an aspartic acid residue at the position of glycan attachment (Liu et al., 2005), we used a monoisotopic mass increment of 0.9848 Da for glycosylation sites in the database search and the following validation procedure using Peptide/Protein Prophet.

A total of 132 proteins were confidently identified with ProteinProphet probability $\geq$ 0.9 (301 unique peptides) to peptide sequences in human IPI database. Of these, 91 proteins (69%) out of them were identified with two or more peptide matches, and 41 proteins (31%) were identified with a single peptide match. A complete list of the identified proteins is provided in the Supplementary Table 1. For the identified proteins, spectral count (total number of peptides) was plotted against peptide count (number of unique peptides; Fig. 1B). Each dot denotes a protein identified in all LC-MS/MS runs. An enlarged view for the proteins with peptide count $\leq$ 5 is represented at the bottom of Fig. 1B. About 78% of the proteins were identified based on 1-2 unique peptides and about 54% were identified based on 1-2 scans. This is clearly due to the effect of N-glycopeptide enrichment since there are only 3 potential glycosylation sites per every protein molecule on average (Apweiler et al., 1999). Our data were compared with those of the Goodlett group, who analyzed the secretomes from prostate and bladder stromal mesenchymal cells (Goo et al., 2009). Their analysis produced a list of 116 glycoproteins for the prostate cells and 84 glycoproteins for the bladder cells; these numbers were similar to the number of proteins we identified with breast cancer cell line Hs578T. While almost all bladder proteins were commonly identified in prostate cells, only 26 bladder proteins and 37 prostate proteins were shared by our breast cells, and 95 proteins were uniquely identified (Fig. 1C). This suggests that secretome constituents are highly dependent on the origin of the cells and that cell line secretome is appropriate for the study of the disease relevant to the cells used (Yi et al., 2002).

Detection of N-glycosylated secretome in the human plasma by western analysis

We next selected target proteins by imposing the following requirements and tested their existence in human plasma. Among the 132 identified proteins, 24 proteins were first picked up that had one or more *N*-glycosylation motifs in matched peptides (64 protein, 48.5%) were in the Secreted Protein Database (SPD; http://spd.cbi.pku.edu.cn/, 85 proteins, 64%), but not yet in the Plasma Proteome Database (PPD; http://www.



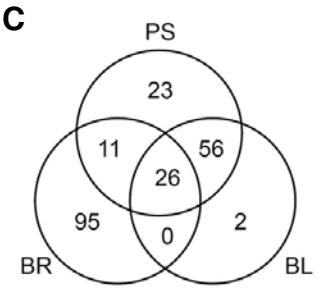


Fig. 1. Analysis of N-glycoprotein enriched from the secretome of Hs578T cells. (A) Workflow for the preparation of secretome, enrichment of glycoproteins, mass spectrometry and data analysis. (B) Spectral count (total number of peptides) as a function of peptide count (number of unique peptides). Each dot denotes a protein identified in all LC-MS/MS runs. An enlarged view for the proteins with a $\leq$ 5 peptide count is shown at the bottom. (C) Glycoproteins of secretomes from prostate and bladder stromal mesenchymal cells (Goo et al., 2009) were combined with the current data. BR, Hs578T breast cell line; PS, prostate cells; BL, bladder cells.

plasmaproteomedatabase.org/, 76 proteins, 58%) (Table 1). The SPD has been set up for proteins that are predicted to be secreted. The majority of secreted proteins have a signal peptide according to the signal hypothesis (Blobel, 2000), and the secreted protein prediction methods have been developed mainly based on the analysis of signal peptides (Chen et al., 2005). The PPD is geared to providing detailed functional annotations of 3,778 distinct proteins based on proteins previously published in the literature as plasma proteins and the identification of multiple peptides from proteins under HUPO's Plasma Proteome Project. The target list was further narrowed down to 13 proteins by considering the availability and specificity of commercial antibodies and by excluding the collagen family. The 13 proteins were versican (VCAN), tyrosine-protein kinase TXK (TXK), insulin-like growth factor binding protein 3 (IGFBP3), N-acetylglucosamine-6-sulfatase (GNS), calumenin (CALU), bone morphogenetic protein 1 (BMP1), FK506-binding protein 10 (FKBP10), urokinase plasminogen activator surface receptor (PLAUR), protein-lysine 6-oxidase (LOX), golgi phosphoprotein 2 (GOLPH2), CD59 glycoprotein (CD59), cathepsin L (CTSL), and prosaposin (PSAP). Although GNS and TXK were not found in human SPD, both of them matched their orthologs in the mouse or rat SPD. The two proteins shared 92% and 83% identity with mouse orthologs. Therefore, we included GNS and TXK in our target list. Interestingly, some of these proteins were well known from their relation to breast or other types of cancer. For example, VCAN plays a role in intercellular signaling and in connecting cells with the extracellular matrix. And VCAN-rich extracellular matrices exert an anti-adhesive effect on the cells, thus facilitating tumor cell migration and invasion (Touab et al., 2002). IGF-binding proteins prolong the half-life of the IGFs and have been shown to either inhibit or stimulate the growth promoting effects of the IGFs on cell culture. Moreover, exogenous IGFBP3 treatment resulted in a significant inhibitory effect on monolayer growth of Hs678T cells (Oh et al., 1993). LOX is responsible for the post-translational oxidative deamination of peptidyl lysine residues in precursors to fibrous collagen and elastin. In addition to cross-linking of extracellular matrix proteins, have a direct role in tumor suppression (Csiszar, 2001). CTSL is a kind of cystein protease that is known to facilitate the invasion and metastasis of tumor cells by degrading the components of basement membrane and extracellular matrix (Goulet et al., 2007; Tsukuba et al., 2000). PSAP interacts with procathepsin D in human breast cancer cells, suggesting a role in tumor invasiveness, metastasis, and the anti-apoptotic effect (Meijer et al., 2009).

Blood plasmas gathered from 6 breast cancer women were pooled and depleted of the top six abundant plasma proteins by immunoaffinity chromatography (Kim et al., 2009), and were then subjected to western analysis after SDS-PAGE. Of the thirteen proteins tested, nine proteins, excluding VCAN, PLAUR, GOLPH2, and CD59, were clearly detected in the plasma (Fig. 2, Table 1). When the plasma sample was treated with PNGase F before SDS-PAGE, eight of the bands detected by western analysis showed increased mobility, suggesting that were glycosylated in plasma (Fig. 2). In contrast, PSAP showed no-change in mobility. According to a previous report, the molecular weight of native human PSAP was estimated to be 66 kDa,

Table 1. Hs578T secretome selected for validation study in blood plasma

No.	Description	Swissprot[a]	Prob[b]	Cov[c]	Num. unique Peps[d]	N-Glycopeptide sequence[e]	SPD[f]	Western analysis[g]	Targeted LC-MS/MS[h]
1	Insulin-like growth factor binding protein 3	P17936	1	20.6	6	KVDYESQSTDTQN@FSSESK VDYESQSTDTQN@FSSESK GLC#VN@ASAVSR	1	1	1
2	N-acetylglucosamine-6-sulfatase	P15586	1	8.9	7	TPMTN@SSIQFLDNAFR TN@SSIQFLDNAFR	2	1	1
3	Urokinase plasminogen activator surface receptor	Q03405	1	6.4	1	GN@STHGC#SSEETFLIDC#R	1	0	
4	Tyrosine-protein kinase TXK	P42681	1	2.3	2	KN@DSGQWYVAER	2	1	1
5	Interleukin-1 receptor accessory protein	Q9NPH3	1	4.6	2	PTLLN@DTGNYTC#M^LR LLN@DTGNYTC#M^LR	1		
6	Prosaposin	P07602	1	5.4	3	TN@STFVQALVEHVK	1	1	1
7	CD59 glycoprotein	P13987	1	12.5	2	TAVN@C#SSDFDAC#LITK	1	0	
8	Cysteine-rich motor neuron 1 protein	Q9NZV1	1	3	3	N@ESC#GGTFGIYGTC#DR PPAC#GELSN@C#TLTGK	1		
9	FK506-binding protein 10	Q96AY3	1	13.7	4	YHYN@GSLM^DGTLFDSSYSR	1	1	1
10	Protein-lysine 6-oxidase	P28300	1	4.3	8	DPGAAVPGAAN@ASAQQPR GAAVPGAAN@ASAQQPR PGAAVPGAAN@ASAQQPR AAVPGAAN@ASAQQPR	1	1	1
11	Biglycan preproprotein variant	P21810	1	4.3	2	M^IEN@GSLSFLPTLR	1		
12	Collagen alpha 2(VI) chain	P12110	1	3.4	4	N@M^TLFSDLVAEK	1		
13	Calumenin	O43852	1	6.2	2	N@ATYGYVLDDPDPDDGFNYK	1	1	0
14	Follistatin-related protein 1	Q12841	1	11	16	FVEQN@ETAINITTYPDQENNK FVEQN@ETAIN@ITTYPDQENNK FVEQNETAIN@ITTYPDQENNK GSN@YSEILDK	1		
15	Prolyl 3-hydroxylase 1	Q7KZR4	1	2	1	VPLQSAHLYYN@VTEK	1		
16	Bone morphogenetic protein 1	P13497	1	7	3	AAVPGN@TSTPSC#QSTNGQPQR	1	1	1
17	Cathepsin L	P07711	1	4.2	2	YSVAN@DTGFVDIPK	1	1	1
18	Alpha-1 type II collagen	P02458	1	1.1	2	LLSTEGSQN@ITYHC#K	1		
19	Low-density lipoprotein receptor-related protein 2	P98164	1	0.4	3	ERTC#AENIC#EQN@CTQLNE	1		
20	Versican core protein	P13611	1	2	2	FEN@QTGFPPPDSR	1	0	
21	PTK7 protein tyrosine kinase 7 isoform a variant	Q13308	1	2.5	2	DDAGN@YTC#IASNGPQGQIR	1		
22	Collagen alpha 1(VI) chain	P12109	1	3.2	2	N@VTAQIC#IDK	1		
23	DKFZP434B0335 protein	Q9UFR6	1	1.4	2	FLNEVVALVPVLN@ETK	1		
24	GolGi phosphoprotein 2	Q8NBJ4	1	3.1	4	AVLVNN@ITTGER	1	0	

[a]Swiss-Prot accession number
[b]ProteinProphet probability
[c]Percent sequence coverage of the matched peptides
[d]Number of identified unique peptides
[e]Sequence of identified N-glycopeptide (C#, alkylated cysteine; M^, oxidized methionine; N@, glycosylated asparagines)
[f]Secreted Protein Database, '1': human database, '2': mouse or rat database.
[g]The proteins were tested for their existence in plasma by western analysis and [h]by targeted LC-MS/MS; '1': tested and confirmed, '0': tested, but not confirmed.

Table 2. The proteotypic peptides used for targeted LC-MS/MS

Name	Sequence[a]	Molecular mass[b]	Target precursor (m/z)	Identified precursor
Insulin-like growth factor binding protein 3	SAGSVESPSVSSTHR	1486.70	744.36 (+2)[c], 496.57 (+3)	
	YGQPLPGYTTK	1223.62	612.82 (+2), 408.88 (+3)	612.82
Calumenin	TFDQLTPEESK	1293.61	1294.62 (+1), 647.81 (+2)	
	HLVYESDQNK	1231.58	1232.59 (+1), 616.80 (+2)	
Cathepsin L	VFQEPLFYEAPR	1494.75	748.38 (+2), 499.26 (+3)	748.38
	GYVTPVK	762.43	763.44 (+1), 382.22 (+2)	
Protein-lysine 6-oxidase	TAGSSGVTAGR	962.48	482.25 (+2), 321.83 (+3)	
	NQGTSDFLPSR	1220.58	611.30 (+2), 407.87 (+3)	611.30
Tyrosine-protein kinase TXK	TQISLSTDEELPEK	1588.78	795.40 (+2), 530.60 (+3)	795.40
	SNLQVVEAISEGFR	1547.79	774.91 (+2), 516.94 (+3)	
N-Acetylglucosamine-6-sulfatase	IQEPNTFPAILR	1397.77	699.89 (+2), 466.93 (+3)	699.89
	SDVLVEYQGEGR	1350.64	676.33 (+2), 451.22 (+3)	
Prosaposin	EIVDSYLPVILDIIK	1728.99	865.50 (+2), 577.34 (+3)	865.50
	QEILAALEK	1013.58	507.80 (+2), 338.87 (+3)	
FK506-binding protein 10	ASPAGGPLEDVVIER	1508.78	755.40 (+2), 503.94 (+3)	
	NTLVAIVVGVGR	1196.72	599.37 (+2), 399.92 (+3)	599.37
Bone morphogenetic protein 1	AAAFLGDIALDEEDLR	1717.85	859.93 (+2), 573.63 (+3)	859.93
	AFQVQQAVDLR	1273.68	637.85 (+2), 425.57 (+3)	

[a]The sequences are shown in single letter codes.
[b]Calculated monoisotopic molecular mass.
[c]Charge states are denoted in parentheses.

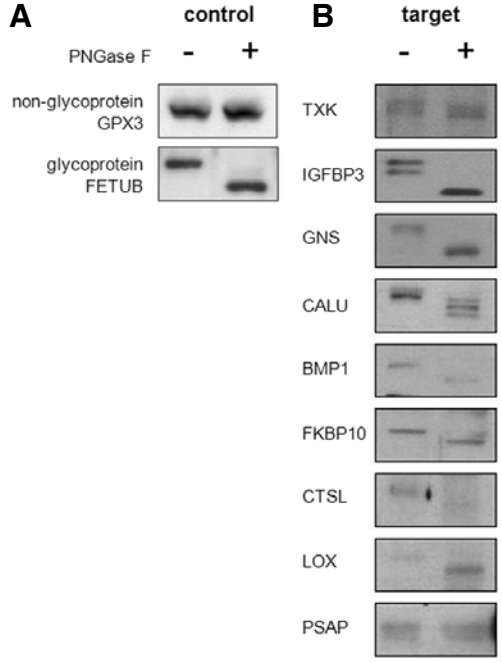


Fig. 2. Western analysis of the identified glycoproteins. Blood plasmas were gathered from six breast cancer patients, pooled, and analyzed by western analysis with the indicated antibodies. The samples were treated with (+)/ without (-) PNGase F before electrophoresis. TXK: tyrosine-protein kinase TXK, IGFBP3: insulin-like growth factor binding protein 3, GNS: N-acetylglucosamine-6-sulfatase, CALU: calumenin, BMP1: bone morphogenetic protein 1, FKBP10: FK506-binding protein 10, CTSL: cathepsin L, LOX: protein-lysine 6-oxidase, PSAP: prosaposin, GPX3: glutathione peroxidase 3, FETUB: fetuin-b.

and deglycosylation of native PSAP yielded a protein with a molecular mass of 54 kDa (Hiraiwa et al., 1993). PASP was migrated as a band above 62 kDa in our experiment. Therefore, it is likely that the PSAP existed in a glycosylated form, but was not deglycosylated by PNGase F, presumably because it was not inaccessible to enzyme. As a control, two plasma proteins, glutathione peroxidase 3 (GPX3) and fetuin-b (FETUB), were tested with PNGase F. FETUB is a glycoprotein while GPX3 is not. As expected, the western band for FETUB shifted after PNGase F treatment, but GPX3 did not show any change in mobility. This suggests that the deglycosylation reaction had been carried out efficiently. The other four proteins (VCAN, PLAUR, GOLPH2, and CD59) that were not detected by western analysis are not likely to be present in plasma. However, other possibilities cannot be ignored. On possibility is that the use of immunoaffinity to remove high-abundant proteins could have resulted in the loss of low-abundant proteins of interest due to non-specific binding (Liotta et al., 2003). Another possibility is that the concentration of these proteins was too low in plasma to be detected with the antibodies used. Nevertheless, if we define plasma proteins as proteins detectable in plasma by currently available techniques, the four proteins should be categorized as non-plasma proteins. Taken together, the results show that nine proteins were definitely detected in blood plasma, which suggests that the low-abundant N-glycoproteins TXK, IGFBP3, GNS, CALU, BMP1, FKBP10, LOX, CTSL, and PSAP should be listed in the PPD.

Detection of N-glycoprotein in human plasma by targeted LC-MS/MS

Detection of target proteins in human plasma was also attempted by LC-MS/MS on in-gel digested samples. The plasma samples were subjected to SDS-PAGE, and the region corresponding to the molecular weight of each protein estimated by western analysis was excised after Coomassie staining. Tryptic peptides were recovered by in-gel digestion and analyzed by C_{18} RPLC-MS/MS (Fig. 3A). During MS/MS, data-dependent acquisition based on a prior survey scan was avoided. Instead, MS/MS were carried out for only the pre-set precursor peptides. We selected two proteotypic peptides for each protein after consulting the GPM database (Table 2). The selected proteo-

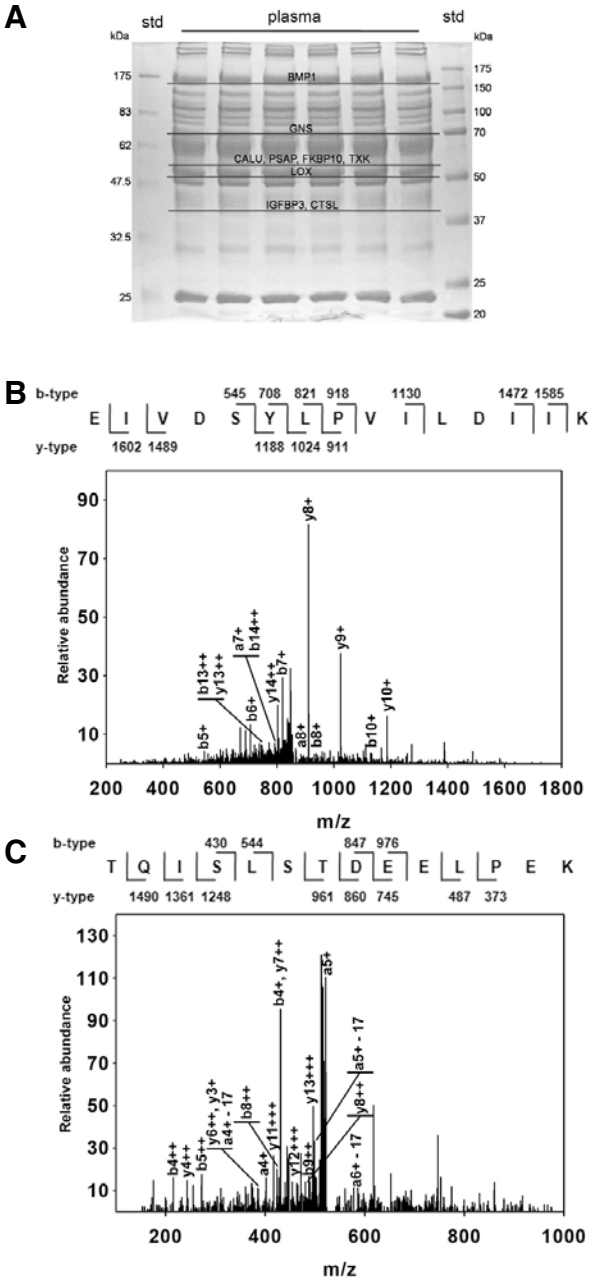


Fig. 3. Verification of glycoprotein in plasma by targeted LC-MS/MS. (A) Pooled plasma was resolved on 10% SDS gels and the gel was stained with Coomassie Brilliant Blue. Solid lines represent the regions corresponding to the molecular weights of the denoted proteins estimated by western analysis. Std: molecular weight standard. (B) MS/MS spectrum of the proteotypic peptide, EIVDSYLPVILDIIK, from prosaposin (PSAP). (C) MS/MS spectrum of the proteotypic peptide, TQISLSTDEELPEK, from tyrosine-protein kinase TXK (TXK).

typic peptides 1) were about 10-16 amino acids, 2) did not contain cysteine, methionine, or tryptophan residues, because these amino acids are high susceptible to oxidation, and 3) included tryptic termini at both sides. The *m/z* values of the proteotypic peptides are listed in Table 2.

Eight proteins, excluding CALU, were confidently identified in blood plasma by targeted LC-MS/MS. All of the identified proteins were detected with precursor ions of charge state two. Two MS/MS spectra supporting the detection of PSAP corresponding to EIVDSYLPVILDIIK and the detection of TXK corresponding to TQISLSTDEELPEK are shown in Figs. 3B and 3C. Tandem mass spectra for the other proteins are shown in Supplementary Fig. 1. The fact that none of these proteins are listed in the PPD implies that these proteins have not been detected so far in serum or plasma by data-dependent mass spectrometry using any type of MS platform. By targeted LC-MS/MS, however, we were able to detect these proteins, and it should be stressed that the mentioned 'target' was acquired from the analysis of a cell line secretome.

CONCLUSION

The proteomic analysis of conditioned media for the discovery of plasma proteins offers several advantages in detecting low abundant proteins over direct analysis of blood because the process enables removal of contamination by circulating immune cells and is not affected by high abundant proteins originating mostly from liver cells. Almost half of the proteins identified from the secretome of Hs578T are currently included in the PPD, and several proteins that were not selected in the current study are known to be secreted by various circulating cells or from contacting tissues as evidenced by the gene ontology database (Supplementary Table 1). Furthermore, the eight proteins we identified from the human plasma provide support for the cell line secretome as a useful source of low abundant serum proteins. Since proteins secreted from a cancer cell line play a central role in pathophysiology of the cells, the proteins are expected to provide not only promising biomarkers but also provide potential targets for new drug development.

Proteins with a relatively low concentration in plasma were clearly detected by targeted LC-MS/MS and western analysis. Our results also provide a platform in the discovery of cancer markers by introduction of a quantitative method such as SILAC (Mann, 2006) or iTRAQ (Ross et al., 2004). Proteins secreted at different levels under various physiological conditions have potentials as serological markers but have to be tested in blood. Antibodies directed against these proteins are usually required for follow-up studies. Despite the considerable effort expended by the Human Protein Atlas, which intends to generate antibodies to all nonredundant human proteins (Guerrier et al., 2008), antibody resources are too limited to provide an antibody against every candidate marker protein. According to our results, it would be worthwhile to make antibodies for proteins detected in cell line secretome in such a translational research pursuing serological biomarkers.

The HUPO has completed an initial large-scale collaborative study to characterize the human serum and plasma proteome. Subsets of 3,778 proteins were identified by mass spectrometry (States et al., 2006). The eight new proteins that we identified, however, have not as yet been identified and therefore are not listed. We suggest that they should be included in the list of proteins in the human plasma proteome based on the results of our indirect proteomics approach.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENT

This study was supported by a grant (FPR08A1-090) from the Functional Proteomics Center of the 21st Century Frontier R&D Program funded by the KOREA Ministry of Education, Science

and Technology, Republic of Korea.

REFERENCES

Amin, S.A., Huang, C.C., Reierstad, S., Lin, Z., Arbieva, Z., Wiley, E., Saborian, H., Haynes, B., Cotterill, H., Dowsett, M., et al. (2006). Paracrine-stimulated gene expression profile favors estradiol production in breast tumors. Mol. Cell. Endocrinol. *253*, 44-55.

Apweiler, R., Hermjakob, H., and Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim. Biophys. Acta *1473*, 4-8.

Bause, E. (1983). Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. Biochem. J. *209*, 331-336.

Blobel, G. (2000). Protein targeting (Nobel lecture). Chembiochem *1*, 86-102.

Chang, J.W., Kang, U.B., Kim, D.H., Yi, J.K., Lee, J.W., Noh, D.Y., Lee, C., and Yu, M.H. (2008). Identification of circulating endorepellin LG3 fragment:Potential use as a serological biomarker for breast cancer. Proteomics Clin. Appl. *2*, 23-32.

Chen, Y., Zhang, Y., Yin, Y., Gao, G., Li, S., Jiang, Y., Gu, X., and Luo, J. (2005). SPD--a web-based secreted protein database. Nucleic Acids Res. *33*, D169-173.

Csiszar, K. (2001). Lysyl oxidases: a novel multifunctional amine oxidase family. Prog. Nucleic Acid Res. Mol. Biol. *70*, 1-32.

Goo, Y.A., Liu, A.Y., Ryu, S., Shaffer, S.A., Malmstrom, L., Page, L., Nguyen, L.T., Doneanu, C.E., and Goodlett, D.R. (2009). Identification of secreted glycoproteins of human prostate and bladder stromal cells by comparative quantitative proteomics. Prostate *69*, 49-61.

Goulet, B., Sansregret, L., Leduy, L., Bogyo, M., Weber, E., Chauhan, S.S., and Nepveu, A. (2007). Increased expression and activity of nuclear cathepsin L in cancer cells suggests a novel mechanism of cell transformation. Mol. Cancer Res. *5*, 899-907.

Guerrier, L., Righetti, P.G., and Boschetti, E. (2008). Reduction of dynamic protein concentration range of biological extracts for the discovery of low-abundance proteins by means of hexapeptide ligand library. Nat. Protoc. *3*, 883-890.

Hakomori, S. (2001). Tumor-associated carbohydrate antigens defining tumor malignancy: basis for development of anti-cancer vaccines. Adv. Exp. Med. Biol. *491*, 369-402.

Hanash, S.M., Pitteri, S.J., and Faca, V.M. (2008). Mining the plasma proteome for cancer biomarkers. Nature *452*, 571-579.

Hiraiwa, M., O'Brien, J.S., Kishimoto, Y., Galdzicka, M., Fluharty, A.L., Ginns, E.I., and Martin, B.M. (1993). Isolation, characterization, and proteolysis of human prosaposin, the precursor of saposins (sphingolipid activator proteins). Arch. Biochem. Biophys. *304*, 110-116.

Huang, L.J., Chen, S.X., Huang, Y., Luo, W.J., Jiang, H.H., Hu, Q.H., Zhang, P.F., and Yi, H. (2006). Proteomics-based identification of secreted protein dihydrodiol dehydrogenase as a novel serum markers of non-small cell lung cancer. Lung Cancer *54*, 87-94.

Keller, A., Eng, J., Zhang, N., Li, X.J., and Aebersold, R. (2005). A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol. Syst. Biol. *1*, 2005 0017.

Kim, Y.J., and Varki, A. (1997). Perspectives on the significance of altered glycosylation of glycoproteins in cancer. Glycoconj. J. *14*, 569-576.

Kim, H.J., Kang, H.J., Lee, H., Lee, S.T., Yu, M.H., Kim, H., and Lee, C. (2009). Identification of S100A8 and S100A9 as serological markers for colorectal cancer. J. Proteome Res. *8*, 1368-1379.

Klee, E.W., Finlay, J.A., McDonald, C., Attewell, J.R., Hebrink, D., Dyer, R., Love, B., Vasmatzis, G., Li, T.M., Beechem, J.M., et al. (2006). Bioinformatics methods for prioritizing serum biomarker candidates. Clin. Chem. *52*, 2162-2164.

Lawlor, K., Nazarian, A., Lacomis, L., Tempst, P., and Villanueva, J. (2009). Pathway-based biomarker search by high-throughput proteomics profiling of secretomes. J. Proteome Res. *8*, 1489-1503.

Lee, H., Yi, E.C., Wen, B., Reily, T.P., Pohl, L., Nelson, S., Aebersold, R., and Goodlett, D.R. (2004). Optimization of reversed-phase microcapillary liquid chromatography for quantitative proteomics. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. *803*, 101-110.

Liotta, L.A., Ferrari, M., and Petricoin, E. (2003). Clinical proteomics: written in blood. Nature *425*, 905.

Liu, T., Qian, W.J., Gritsenko, M.A., Camp, D.G., 2nd, Monroe, M.E., Moore, R.J., and Smith, R.D. (2005). Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. J. Proteome Res. *4*, 2070-2080.

Mann, M. (2006). Functional and quantitative proteomics using SILAC. Nat. Rev. Mol. Cell Biol. 7, 952-958.

Meijer, D., Jansen, M.P., Look, M.P., Ruigrok-Ritstier, K., van Staveren, I.L., Sieuwerts, A.M., van Agthoven, T., Foekens, J.A., Dorssers, L.C., and Berns, E.M. (2009). TSC22D1 and PSAP predict clinical outcome of tamoxifen treatment in patients with recurrent breast cancer. Breast Cancer Res. Treat. *113*, 253-260.

Oh, Y., Muller, H.L., Lamson, G., and Rosenfeld, R.G. (1993). Insulin-like growth factor (IGF)-independent action of IGF-binding protein-3 in Hs578T human breast cancer cells. Cell surface binding and growth inhibition. J. Biol. Chem. *268*, 14964-14971.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., et al. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics *3*, 1154-1169.

Roth, J. (2002). Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. Chem. Rev. *102*, 285-303.

Siriwardana, G., Bradford, A., Coy, D., and Zeitler, P. (2006). Autocrine/paracrine regulation of breast cancer cell proliferation by growth hormone releasing hormone via Ras, Raf, and mitogen-activated protein kinase. Mol. Endocrinol. *20*, 2010-2019.

States, D.J., Omenn, G.S., Blackwell, T.W., Fermin, D., Eng, J., Speicher, D.W., and Hanash, S.M. (2006). Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. Nat. Biotechnol. *24*, 333-338.

Touab, M., Villena, J., Barranco, C., Arumi-Uria, M., and Bassols, A. (2002). Versican is differentially expressed in human melanoma and may play a role in tumor development. Am. J. Pathol. *160*, 549-557.

Tsukuba, T., Okamoto, K., Yasuda, Y., Morikawa, W., Nakanishi, H., and Yamamoto, K. (2000). New functional aspects of cathepsin D and cathepsin E. Mol. Cells *10*, 601-611.

Wu, C.C., Chien, K.Y., Tsang, N.M., Chang, K.P., Hao, S.P., Tsao, C.H., Chang, Y.S., and Yu, J.S. (2005). Cancer cell-secreted proteomes as a basis for searching potential tumor markers: naso pharyngeal carcinoma as a model. Proteomics *5*, 3173-3182.

Yi, E.C., Marelli, M., Lee, H., Purvine, S.O., Aebersold, R., Aitchison, J.D., and Goodlett, D.R. (2002). Approaching complete peroxisome characterization by gas-phase fractionation. Electrophoresis *23*, 3205-3216.

Zhang, H., Li, X.J., Martin, D.B., and Aebersold, R. (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat. Biotechnol. *21*, 660-666.

Zhang, H., Yi, E.C., Li, X.J., Mallick, P., Kelly-Spratt, K.S., Masselon, C.D., Camp, D.G., 2nd, Smith, R.D., Kemp, C.J., and Aebersold, R. (2005). High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. Mol. Cell. Proteomics *4*, 144-155.