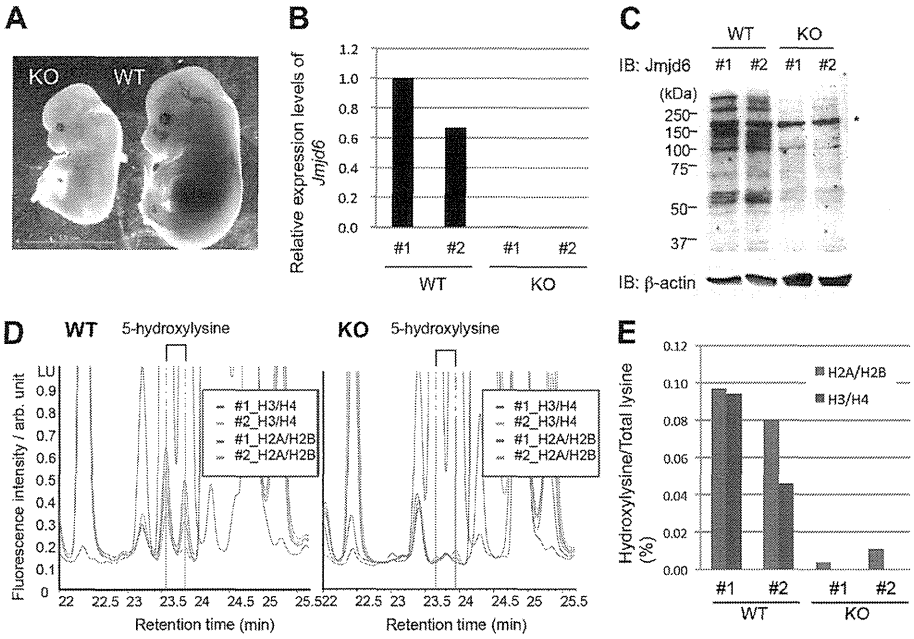


FIGURE 4. **JMJD6 hydroxylates histones H2A/H2B and H3/H4 in mouse embryos.** *A*, a representative image of JMJD6 knock-out and wild-type E14.5 embryos. *B*, *JMJD6* knock-out was confirmed by qRT-PCR. GAPDH was used as an internal control. *C*, JMJD6 knock-out was confirmed by Western blotting. The *asterisk* indicates a nonspecific band. β-Actin was used as a loading control. *D*, result of amino acid composition analysis of histones derived from two Jmjd6 wild-type (*left*) and knock-out (*right*) E14.5 embryos. *E*, % of 5-hydroxylysine in total lysine of histones H2A/H2B (*blue*) and H3/H4 (*red*) was calculated from the HPLC data (*D*). *IB*, immunoblot; *arb. unit*, arbitrary units.

(nM) of the methylated substrate was calculated based on the basis of radioactivity.

RESULTS

JMJD6 Effectively Hydroxylates Histone Lysyl Residues in Vitro—During screening of UHRF1-interacting proteins, we identified JMJD6 as a novel binding partner of UHRF1 (data not shown). Because UHRF1 recognizes hemimethylated DNA and histone modifications, we assumed that JMJD6 might be recruited by UHRF1 to nucleosomes and modify histone lysyl residues. *In vitro* experiments showed that recombinant GST-JMJD6 possessed the ability to bind to histone $H3_{1-20}$ tail and histone H4 (Fig. 1, *A* and *B*) and hydroxylate multiple lysyl residues in the N-terminal tails of histone $H3_{1-20}$ and $H4_{1-30}$, which was detected as of 16, 32, or 48 Da shifts by MS analysis (Fig. 1, *C* and *D*); subsequent MS/MS analysis revealed that JMJD6 mediates monohydroxylation of lysyl residues. As indicated by Webby *et al.* (1), JMJD6 preferentially hydroxylated lysyl residues in the basic peptides, and no apparent sequence preference was observed *in vitro* (data not shown).

Next, we established a sensitive hydroxylysine detection method based on amino acid composition analysis as an alternative to the MS-based method. For amino acid composition analysis, we briefly hydrolyzed peptides or proteins with HCl and separated each amino acid residue by reversed phase HPLC to detect 5-hydroxylysine. To evaluate this method, we first performed reversed phase HPLC using simplicial synthetic 2*S*,5*R*-hydroxylysine and synthetic racemic mixture of 5-hydroxylysine composed of 2*S*,5*S* (*SS*)-, 2*R*,5*R* (*RR*)-, 2*R*,5*S* (*RS*)-, and 2*S*,5*R* (*SR*)-stereoisomers (Fig. 2*A*). We detected two peaks corresponding to *SS*/*RR*- and *RS*/*SR*-hydroxylysine by analyzing these synthetic 5-hydroxylysines without HCl treatment (Fig. 2*A*). After HCl treatment of these synthetic 5-hydroxylysines, another peak was appeared (Fig. 2*A*, *arrow*). This peak possibly corresponds to a lactone derivative, 3-amino-6-(aminomethyl)oxan-2-one, generated by dehydration condensation between C5 hydroxyl group and carboxyl group, which is described in a previous report (11). Next, we evaluated the method using unmodified $H4_{1-23}$ peptides and 5-hydroxylysine containing $H4_{1-23}$ peptides in which all the lysines at positions 5, 8, 12, and 20 were substituted with 5-hydroxylysines. After hydrolysis of these peptides, we detected two peaks corresponding to *SS*/*RR*- and *RS*/*SR*-hydroxylysine only in the 5-hydroxylysine containing peptides but not in the unmodified peptides (Fig. 2, *B* and *C*). We also detected the peak of the possible lactone derivative in the 5-hydroxylysine containing peptides by reversed phase HPLC performed in the same day of hydrolysis, but the peak disappeared in the next day of hydrolysis, indicating that the derivative is unstable. Because quantification of the derivative is technically difficult, we only quantified *SS*/*RR*- and *RS*/*SR*-hydroxylysine.

Using this method, we analyzed $H3_{21-40}$ and $H4_{1-23}$ peptides treated with or without recombinant GST-JMJD6. First, we confirmed hydroxylation of the peptides by GST-JMJD6 by MS analysis (Fig. 3*A*). Then, the peptides were separated from the enzyme reaction mixture, by reversed phase HPLC. The separated peptides were treated with HCl, and each amino acid residue was separated by reversed phase HPLC (Fig. 3, *B* and *C*). Comparison of the chromatograph between amino acids derived from the JMJD6-treated and -untreated peptides identified two additional peaks in the peptides treated with JMJD6, which are matched with the standard synthetic 5-hydroxylysine (Fig. 3, *B* and *C*).

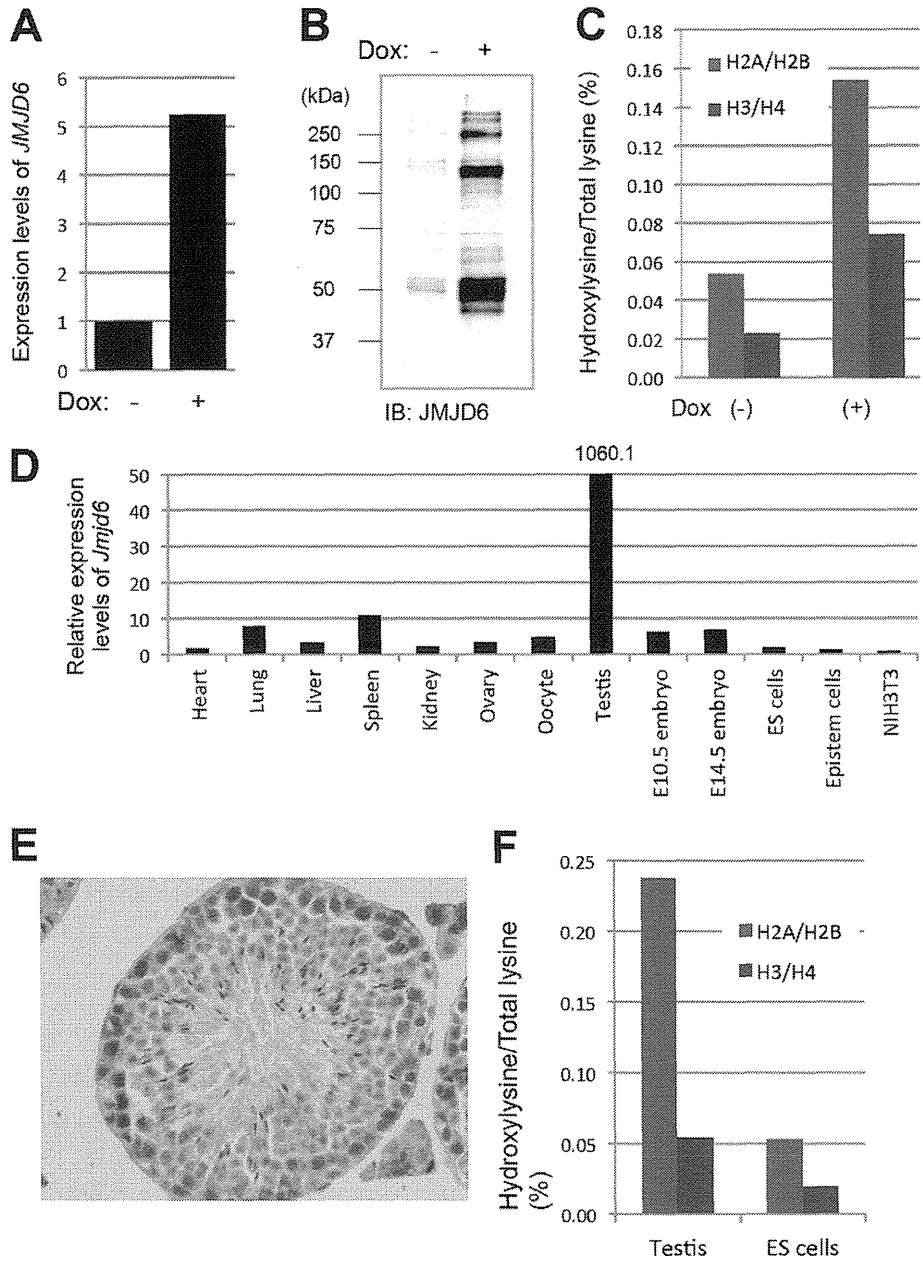


FIGURE 5. **Amount of 5-hydroxylysine in JMJD6 overexpressed HEK 293 cells, mouse testis, and ES cells.** *A*, expression levels of *JMJD6* in Dox-inducible JMJD6 stable cells were examined by qRT-PCR before and after 48-h Dox induction. *B*, induction of JMJD6 by Dox in the cells was confirmed by Western blotting. *C*, amino acid composition analysis of histones derived from Dox-inducible JMJD6 stable cell lines. The *blue* and *red bars* indicate % of 5-hydroxylysine in the total lysine of histone H2A/H2B and in the H3/H4, respectively. *D*, relative expression levels of *Jmjd6* in various mouse tissues and cells were examined by qRT-PCR. *E*, expression of Jmjd6 in a 6-month-old mouse testis was examined by immunohistochemistry. *F*, amino acid composition analysis of histones derived from 6-month-old mouse testis and J1 ES cells. The *blue* and *red bars* indicate % of 5-hydroxylysine in the total lysine of histone H2A/H2B and in the H3/H4, respectively.

JMJD6 Hydroxylates Histone Lysyl Residues in Vivo—To investigate histone lysyl hydroxylation *in vivo*, we performed the amino acid composition analysis for analyzing a mixture of histone H2A/H2B and a mixture of histone H3/H4 proteins isolated from two JMJD6 wild-type and two JMJD6 knock-out whole embryos at E14.5 (Fig. 4*A*). JMJD6 knock-out was confirmed by qRT-PCR and Western blotting (Fig. 4, *B* and *C*). The results showed that 0.097 and 0.080% of total lysyl residues in histone H2A/H2B and 0.094 and 0.046% of those in histone H3/H4 were 5-hydroxylated in each of the two JMJD6 wild-type mice (Fig. 4, *D* and *E*), whereas 0.004 and 0.011% of total lysyl residues in histone H2A/H2B and 0.000 and 0.000% of those in histone H3/H4 were 5-hydroxylated in each of the two JMJD6 knock-out mice (Fig. 4, *D* and *E*), indicating that JMJD6 hydroxylates histone lysyl residues *in vivo*.

We also generated Dox-inducible JMJD6 stable HEK293 cells. JMJD6 induction by Dox was confirmed by qRT-PCR and Western blotting (Fig. 5, *A* and *B*) and increased 5-hydroxylation levels of histone lysyl residues (Fig. 5*C*). In addition, we purified histones from a 6-month-old JMJD6 wild-type mouse

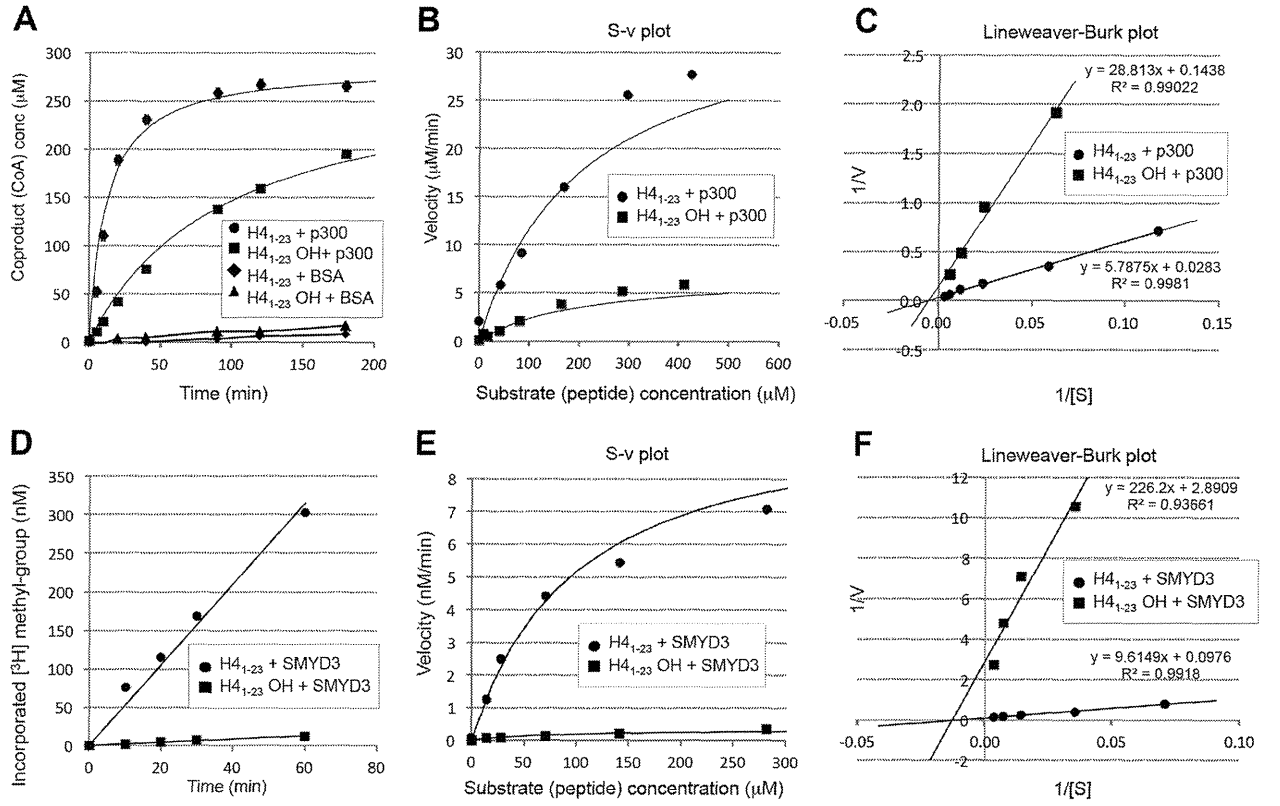


FIGURE 6. **5-Hydroxylation of lysyl residue impairs *N*-acetylation and *N*-methylation *in vitro*.** *A–C*, the *in vitro* colorimetric HAT assay was performed using a fixed amount of p300 (0.44 μM) and control $H4_{1-23}$ peptides (●) or 5-hydroxylysine-containing peptides ($H4_{1-23}$ OH, ■). BSA was used as a negative control (◆, ▲). After the reactions, absorbance (405 nm) of the coproduct (CoA) was measured. *A*, reactions were terminated at the indicated time points, and the concentration of CoA was calculated on the basis of a standard curve that was generated from β2-mercaptoethanol. *B* and *E*, substrate concentration-velocity (s-v) plot. *C* and *F*, Lineweaver-Burk plot. The *vertical axis* is 1/velocity [v], and the *horizontal axis* is 1/substrate concentration [S]. *D–F*, the *in vitro* histone methyltransferase assay was performed using a fixed amount of SMYD3 (1 μM) and control $H4_{1-23}$ peptides (●) or 5-hydroxylysine-containing peptides ($H4_{1-23}$OH, ■). AdoMet was used as a methyl donor. After the reactions, radioactivity (cpm) of the ^{3}H-methylated substrates was measured. The concentration of incorporated ^{3}H-methyl groups (nM) was calculated based on the basis of radioactivity (1 cpm was 0.02563 nM in the reaction). *D*, reactions were performed with fixed amounts of the peptides (141 μM) and terminated at the indicated time points.

testis, which expressed JMJD6 at the highest level among various tissues and cells (Fig. 5, *D* and *E*). In the testis, 0.238 and 0.054% of total lysyl residues in histone H2A/H2B and H3/H4, respectively, were 5-hydroxylated (Fig. 5*F*). In the mouse J1 ES cells, 0.053 and 0.020% of total lysyl residues in histone H2A/H2B and H3/H4, respectively, were 5-hydroxylated (Fig. 5*F*).

5-Hydroxylation Prevents N-Acetylation and N-Methylation of Histone Lysyl Residues in Vitro—Because lysyl residues in histone tails are often subjected to *N*-acetylation and *N*-methylation, we examined whether 5-hydroxylation of lysyl residues affects modifications at the ϵ-amino groups. First, we examined the effect of lysyl 5-hydroxylation on histone H4 *N*-acetylation by p300, which catalyzes *N*-acetylation of the ϵ-amino group of lysyl residues, including histone H4K5 and H4K8, through its HAT domain (12). Kinetic analysis using the unmodified and the 5-hydroxylysine containing $H4_{1-23}$ peptides in which all the lysines were substituted with 5-hydroxylysines as substrates revealed that 5-hydroxylation largely interfered with the HAT activity of p300 *in vitro* (Fig. 6, *A–C*). Lineweaver-Burk plot analysis was performed to calculate the maximum velocity (V_{max}) and Michaelis constant (K_m) values (Table 1; Fig. 6*C*, $R^2 = 0.9981$ and 0.9902). V_{max} of the reactions in which p300 acetylated the 5-hydroxylysine-containing peptides ($H4_{1-23}$OH) was 5-fold less than that of the control peptides (6.95 ± 0.45 and 35.34 ± 1.65 μM/min, respectively), whereas K_m of the two reactions was quite similar (204.51 ± 10.12 and 200.37 ± 14.32 μM, respectively), indicating that 5-hydroxylation does not inhibit binding of lysyl residues to p300 but reduces the catalytic efficiency.

TABLE 1

Effect of 5-hydroxylation on *N*-acetylation of ϵ-amino group of lysyl residues

Lineweaver-Burk plots were used for estimation of the kinetic constants, V_{max}, and K_m. R^2 is the determination coefficient (see Fig. 6*C*).

	V_{max}	K_m
	μM/min	*μM*
$H4_{1-23}$ + p300	35.34 ± 1.65 (R^2 = 0.9981)	204.51 ± 10.12
$H4_{1-23}$ OH + p300	6.95 ± 0.45 (R^2 = 0.9902)	200.37 ± 14.32

TABLE 2

Effect of 5-hydroxylation on *N*-methylation of ϵ-amino group of lysyl residues

Lineweaver-Burk plots were used for estimation of the kinetic constants, V_{max}, and K_m. R^2 is the determination coefficient (see Fig. 6*F*).

	V_{max}	K_m
	nM/min	
$H4_{1-23}$ + SMYD3	10.90 ± 0.92 (R^2 = 0.9918)	80.63 ± 16.26 μM
$H4_{1-23}$ OH + SMYD3	0.48 ± 0.26 (R^2 = 0.9366)	75.26 ± 9.60 μM

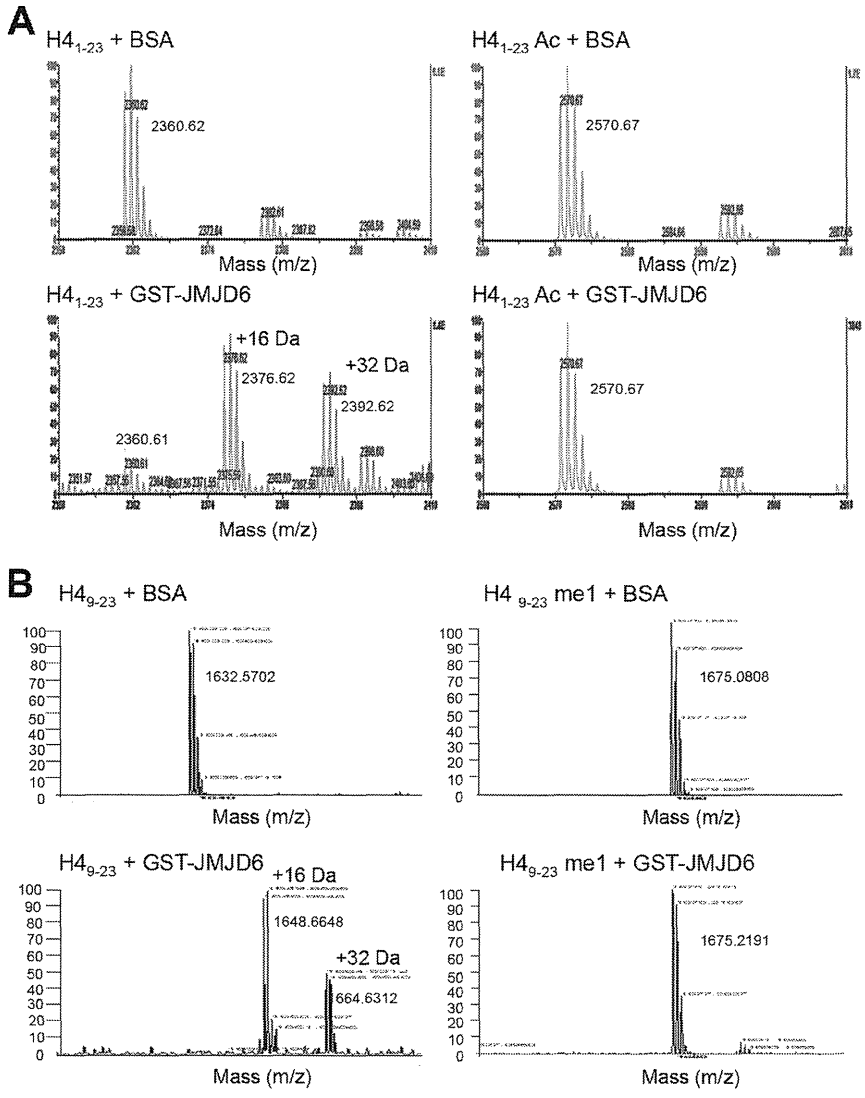


FIGURE 7. ***N*-Acetylation and *N*-methylation of lysyl residues impairs 5-hydroxylation by JMJD6 *in vitro*.** *A*, the *in vitro* hydroxylation assay was performed using GST-JMJD6 (10 μM) and 85 μM of control $H4_{1-23}$ peptides or *N*-acetyl-lysine-containing peptides. *B*, the *in vitro* hydroxylation assay was performed using GST-JMJD6 (10 μM) and 85 μM control $H4_{1-23}$ peptides or *N*-monomethyl-lysine-containing peptides. BSA was used as a negative control. 5-Hydroxylation by JMJD6 was detected by MS analysis.

We also examined the effect of lysyl 5-hydroxylation on the histone methyltransferase activity of SMYD3, which catalyzes lysyl *N*-methylation of histone H3 (13) and also H4 (data not shown) through its SET (su(var) 3–9 enhancer-of-zeste trithorax) domain by the histone methyltransferase assay. The results showed that 5-hydroxylation at lysyl residues almost completely inhibited *N*-methylation catalyzed by SMYD3 (Table 2 and Fig. 6, *D–F*); V_{max} values of the reactions with the control peptides ($H4_{1-23}$) and the 5-hydroxylysine-containing peptides ($H4_{1-23}OH$) as substrates were 10.90 ± 0.92 and 0.48 ± 0.26 nM/min, respectively. Similar to the HAT assay, K_m values of the two reactions were ~80.63 ± 16.26 and 75.26 ± 9.60 μM, respectively.

Subsequently, we performed reciprocal experiments using $H4_{1-23}$ or $H4_{9-23}$ peptides, in which all the lysines are either unmodified, *N*-acetylated, or *N*-monomethylated. JMJD6 effectively hydroxylated the control peptides (Fig. 7, *A* and *B*, *left panels*); however, *N*-acetylation and *N*-monomethylation at the ϵ-amino group of the lysines completely blocked 5-hydroxylation by JMJD6 (Fig. 7, *A* and *B*, *right panels*).

DISCUSSION

We found a novel histone modification, 5-hydroxylation, by JMJD6. JMJD6 reportedly hydroxylates a splicing factor, U2AF65 (1). That study and another report (1, 14) stated that evidence of histone lysyl hydroxylation was not found by MS-based analysis *in vivo*. In the present study, we developed an alternative method, amino acid composition analysis, to detect 5-hydroxylation of histone lysyl residues. As reported previously, we have not detected clear evidence of 5-hydroxylation of histone lysyl residues by MS-based analysis. We think that there are several causes for this. 1) The amount of 5-hydroxylysine is too small to detect by MS-based analysis. 2) Artificial methionine oxidation during preparation of samples for MS analysis

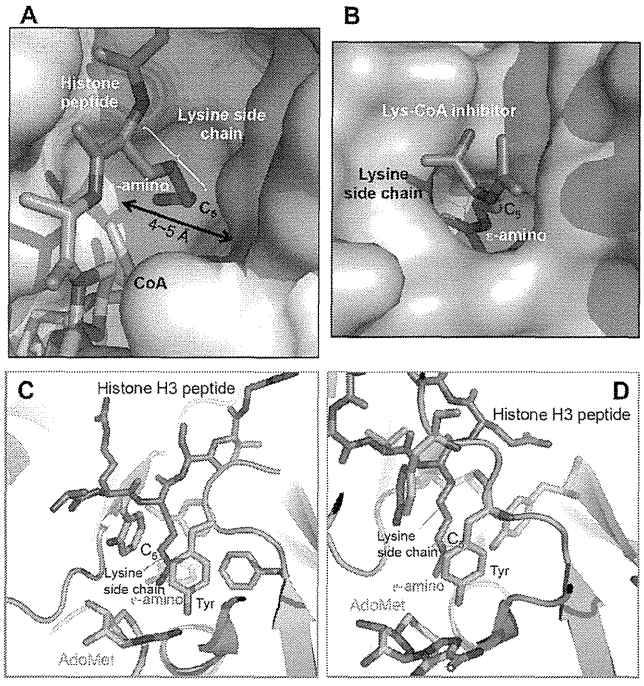


FIGURE 8. **Structure around the active site of HAT domains and SET domains, a lysine side chain, and an *S*-adenosyl methionine (*AdoMet*).** *A*, spatial localization among the HAT domain of GCN5 (*gray*), a lysine side chain (*magenta*), and CoA (*cyan*) (Protein Data Bank code 1QSN). The 5-hydroxyl group may locate close to acetyl-CoA, indicating that this may work as a steric barrier and prevent effective *N*-acetylation by the catalytic domain. *B*, structure of the HAT domain of p300 (*gray*) in complex with inhibitor, Lys-CoA (*magenta*) (Protein Data Bank code 3BIY). The 5-hydroxyl group may restrict the conformation of lysine side chain in the catalytic pocket of p300. The figures were prepared by using program PyMOL. *C*, spatial localization among the SET domain of *N. crassa* Dim-5 (*green*), a histone H3 peptide (*magenta*), and AdoMet (*cyan*) (Protein Data Bank code 1PEG). *D*, spatial association among the SET domain of human SETD7/9 (*green*), a histone H3 peptide containing monomethylated Lys-4 (*magenta*), and AdoMet (*cyan*) (Protein Data Bank code 1O9S). A side chain of a lysine residue locates in tightly hydrophobic pocket of the SET domains (*C* and *D*). Hydroxylation at position C_5 of the chain may prevent a lysine side chain to locate in the pocket, causing inhibition of *N*-methylation by SET domains.

makes detection of 5-hydroxylysine difficult. 3) 5-Hydroxylysine could be an intermediate form as it is in collagen, and a further unknown modification(s) such as glycosylation could be added; the final product of collagen is glucosylgalactosyl hydroxylysine (4). The collagen hydroxylase, PLOD3, possesses galactosyltransferase and glucosyltransferase activities. Unlike PLOD3, JMJD6 does not appear to possess any other enzymatic activities by domain search; therefore, it is difficult to assume possible further modification(s) by its protein structure. We may have been able to detect 5-hydroxylation in histone lysyl residues by amino acid composition analysis but not by the MS-based analysis because many modifications such as glycosylation or galactosylation could be removed during the hydrolysis process of amino acid composition analysis. By this analysis, we detected both *SS*/*RR*- and *SR*/*RS*-hydroxylysine in JMJD6-treated histone peptides and also in JMJD6 wild-type E14.5 embryos, ES cells, and the Dox-inducible JMJD6 stable HEK293 cells. Because the relationship between *RR* and *RS* and also between *SS* and *RS* is a diastereomer, we were able to distinguish them. However, because a relationship between *SS* and *RR*, and also between *RS* and *SR* is an enantiomer, we were not able to separate them by this method. Despite this, these two peaks are most likely *SS*- and *RS*-hydroxylysine because JMJD6 is reported to generate *SS*-hydroxylysine (11). The *RS*-hydroxylysine could be generated from *SS*-hydroxylysine through the lactone derivative, 3-amino-6-(aminomethyl)oxan-2-one, which is unstable and difficult to be quantified. Because of this difficulty, we only quantified *SS*/*RR*- and *RS*/*SR*-hydroxylysine in this report. Therefore, actual quantity of 5-hydroxylysine in the samples examined here could be a little higher.

Because we detected 5-hydroxylysines in the UHRF1 KO ES cells (data not shown), UHRF1 is not required for 5-hydroxylation of histone lysyl residues by JMJD6. Therefore, biological significance of the interaction between UHRF1 and JMJD6 remains unclear. Further analysis is also required to determine the biological significance of 5-hydroxylation of histone lysyl residues. *In vitro* experiments suggest that 5-hydroxylation can inhibit *N*-acetylation and *N*-methylation by p300 and SMYD3. The active site structure of the p300 and general control of amino acid synthesis 5 (GCN5) HAT domains showed that the side chain of the 5-hydroxylysine can invade the catalytic pocket; however, the 5-hydroxyl group may disturb active form formation of the substrate (Fig. 8, *A* and *B*). The catalytic site of SET domains of *Neurospora crassa* Dim-5 and human SETD7, which are structurally similar to SMYD3 (15), suggested that the side chain of 5-hydroxylysine can invade the catalytic pocket; however, the 5-hydroxyl group may disturb an active form formation of the substrate (16, 17) similar to that in HAT domains (Fig. 8, *C* and *D*). Therefore, 5-hydroxylation could be important in the context of the histone code. It is known that histones H2A and H2B move more dynamically between the nucleosome and nucleoplasm. 5-Hydroxylation of these histones may have some effects for the movement because the modification was detected more in histones H2A/H2B than in histones H3/H4. The expression pattern of JMJD6 is also interesting. JMJD6 may play important role(s) in the testis, such as a role in histone-protamine exchange. We believe that our present finding provides a novel insight into epigenetic regulations of gene transcription and/or chromosomal rearrangement.

Acknowledgments—We thank Dr. Haruhiko Koseki and Dr. Jafar Sharif for providing us UHRF1 KO ES cells; Professor Shoji Tajima, Dr. Isao Suetake, Dr. Yoichi Shinkai, Dr. Kenji Ichiyanagi, Dr. Fumiyuki Sanematsu, and Dr. Hyun-Soo Cho for useful advice; and Yuichi Mishima and Dr. Atsuhiko Toyama for technical assistance.

REFERENCES

1. Webby, C. J., Wolf, A., Gromak, N., Dreger, M., Kramer, H., Kessler, B., Nielsen, M. L., Schmitz, C., Butler, D. S., Yates, J. R., 3rd, Delahunty, C. M., Hahn, P., Lengeling, A., Mann, M., Proudfoot, N. J., Schofield, C. J., and Böttger, A. (2009) Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science* **325,** 90–93
2. Hong, X., Zang, J., White, J., Wang, C., Pan, C. H., Zhao, R., Murphy, R. C., Dai, S., Henson, P., Kappler, J. W., Hagman, J., and Zhang, G. (2010) Interaction of JMJD6 with single-stranded RNA. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 14568–14572
3. Loenarz, C., and Schofield, C. J. (2008) Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* **4,** 152–156
4. Myllylä, R., Wang, C., Heikkinen, J., Juffer, A., Lampela, O., Risteli, M., Ruotsalainen, H., Salo, A., and Sipilä, L. (2007) Expanding the lysyl hydrox-

ylase toolbox: new insights into the localization and activities of lysyl hydroxylase 3 (LH3). *J. Cell. Physiol.* **212,** 323–329
5. Shi, Y., and Whetstine, J. R. (2007) Dynamic regulation of histone lysine methylation by demethylases. *Mol. Cell* **25,** 1–14
6. Kunisaki, Y., Masuko, S., Noda, M., Inayoshi, A., Sanui, T., Harada, M., Sasazuki, T., and Fukui, Y. (2004) Defective fetal liver erythropoiesis and T lymphopoiesis in mice lacking the phosphatidylserine receptor. *Blood* **103,** 3362–3364
7. Böse, J., Gruber, A. D., Helming, L., Schiebe, S., Wegener, I., Hafner, M., Beales, M., Köntgen, F., and Lengeling, A. (2004) The phosphatidylserine receptor has essential functions during embryogenesis but not in apoptotic cell removal. *J. Biol.* **3,** 15
8. Unoki, M., Brunet, J., and Mousli, M. (2009) Drug discovery targeting epigenetic codes: the great potential of UHRF1, which links DNA methylation and histone modifications, as a drug target in cancers and toxoplasmosis. *Biochem. Pharmacol.* **78,** 1279–1288
9. Masuda, A., and Dohmae, N. (2010) Automated Protein Hydrolysis Delivering Sample to a Solid Acid Catalyst for Amino Acid Analysis. *Anal. Chem.* **82,** 8939–8945
10. Strahl, B. D., Ohba, R., Cook, R. G., and Allis, C. D. (1999) Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 14967–14972
11. Mantri, M., Loik, N. D., Hamed, R. B., Claridge, T. D., McCullagh, J. S., and Schofield, C. J. (2011) The 2-oxoglutarate-dependent oxygenase JMJD6 catalyses oxidation of lysine residues to give 5S-hydroxylysine residues. *Chembiochem.* **12,** 531–534
12. Schiltz, R. L., Mizzen, C. A., Vassilev, A., Cook, R. G., Allis, C. D., and Nakatani, Y. (1999) Overlapping but distinct patterns of histone acetylation by the human coactivators p300 and PCAF within nucleosomal substrates. *J. Biol. Chem.* **274,** 1189–1192
13. Hamamoto, R., Furukawa, Y., Morita, M., Iimura, Y., Silva, F. P., Li, M., Yagyu, R., and Nakamura, Y. (2004) SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat. Cell Biol.* **6,** 731–740
14. Han, G., Li, J., Wang, Y., Li, X., Mao, H., Liu, Y., and Chen, C. D. (2012) The hydroxylation activity of Jmjd6 is required for its homo-oligomerization. *J. Cell. Biochem.* **113,** 1663–1670
15. Dillon, S. C., Zhang, X., Trievel, R. C., and Cheng, X. (2005) The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol.* **6,** 227
16. Zhang, X., Tamaru, H., Khan, S. I., Horton, J. R., Keefe, L. J., Selker, E. U., and Cheng, X. (2002) Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* **111,** 117–127
17. Subramanian, K., Jia, D., Kapoor-Vazirani, P., Powell, D. R., Collins, R. E., Sharma, D., Peng, J., Cheng, X., and Vertino, P. M. (2008) Regulation of estrogen receptor α by the SET7 lysine methyltransferase. *Mol. Cell* **30,** 336–347

Journal of proteome research

Article

pubs.acs.org/jpr

Plasma Low-Molecular-Weight Proteome Profiling Identified Neuropeptide-Y as a Prostate Cancer Biomarker Polypeptide

Koji Ueda,*[,†] Ayako Tatsuguchi,[†] Naomi Saichi,[†] Atsuhiko Toyama,[‡] Kenji Tamura,[§] Mutsuo Furihata,[∥] Ryo Takata,[⊥] Shusuke Akamatsu,[#] Masahiro Igarashi,[▽] Masato Nakayama,[○] Taka-Aki Sato,[‡] Osamu Ogawa,[#] Tomoaki Fujioka,[⊥] Taro Shuin,[§] Yusuke Nakamura,[◆] and Hidewaki Nakagawa[†]

[†]Laboratory for Biomarker Development, Center for Genomic Medicine, RIKEN, General Research Building 6F, Institute of Medical Science, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

[‡]Life Science Research Center, Shimadzu Corporation, 1-3, Nishikicho, Kanda, Chiyoda-ku, Tokyo 101-8448, Japan

[§]Department of Urology, Kochi Medical School Hospital, Oko-cho Kohasu, Nankoku-shi, Kochi 783-8505, Japan

[∥]Department of Pathology, Kochi Medical School, Oko-cho Kohasu, Nankoku-shi, Kochi 783-8505, Japan

[⊥]Department of Urology, Iwate Medical University, 19-1 Uchimaru, Morioka, Iwate Prefecture 020-0023, Japan

[#]Department of Urology, Graduate School of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

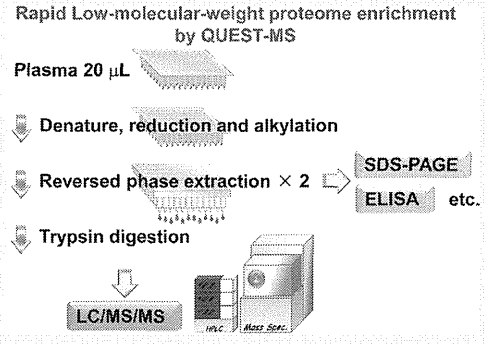
[▽]Endoscopy Division, Gastrointestinal Center, Cancer Institute Hospital, 3-8-31, Ariake, Koto, Tokyo 135-8550, Japan

[○]Toppan Printing Co., Ltd., 1-5-1, Taito, Taito-ku, Tokyo 110-8560, Japan

[◆]Section of Hematology/Oncology, Department of Medicine Faculty, The University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois, United States

Supporting Information

ABSTRACT: In prostate cancer diagnosis, PSA test has greatly contributed to the early detection of prostate cancer; however, expanding overdiagnosis and unnecessary biopsies have emerged as serious issues. To explore plasma biomarkers complementing the specificity of PSA test, we developed a unique proteomic technology QUEST-MS (Quick Enrichment of Small Targets for Mass Spectrometry). The QUEST-MS method based on 96-well formatted sequential reversed-phase chromatography allowing efficient enrichment of <20 kDa proteins quickly and reproducibly. Plasma from 24 healthy controls, 19 benign prostate hypertrophy patients, and 73 prostate cancer patients were purified with QUEST-MS and analyzed by LC/MS/MS. Among 153 057 nonredundant peptides, 189 peptides showed prostate cancer specific detection pattern, which included a neurotransmitter polypeptide neuropeptide-Y (NPY). We further validated the screening results by targeted multiple reaction monitoring technology using independent sample set (n = 110). The ROC curve analysis revealed that logistic regression-based combination of NPY, and PSA showed 81.5% sensitivity and 82.2% specificity for prostate cancer diagnosis. Thus QUEST-MS technology allowed comprehensive and high-throughput profiling of plasma polypeptides and had potential to effectively uncover very low abundant tumor-derived small molecules, such as neurotransmitters, peptide hormones, or cytokines.



KEYWORDS: *low molecular weight, Biomarker, prostate cancer, plasma, PSA, mass spectrometry, label-free quantification*

INTRODUCTION

Prostate-specific antigen (PSA), also known as kallikrein-3, was discovered in 1969[1] and had been recognized as the best diagnostic tool for prostate cancer since the Food and Drug Administration (FDA) approved the PSA test in 1994.[2] Indeed, at least 30 million men over 50 years old undergo PSA test in the United States in a year. However, in October 2011, the U.S. Preventive Services Task Force (USPSTF) published statistical evidence about clinical outcomes of prostate cancer and urged not to continue PSA test against healthy males anymore.[3] This decision was simply based on the concept that the benefit of PSA test to overall survival rate of prostate cancer patients did not worth the risk for expanding invasive prostate biopsy cases and medical costs. It was estimated that 5.2 million U.S. dollars would be spent for PSA screening to prevent one death from prostate cancer.[4] Therefore, a new biomarker set that can efficiently improve the poor specificity of PSA is urgently required for the reduction of risks above derived from overdiagnosis of prostate cancer.

Received: June 12, 2013
Published: August 30, 2013

dx.doi.org/10.1021/pr400547s | J. Proteome Res. 2013, 12, 4497−4506

Table 1. 116 Plasma Samples Used for Biomarker Screening

group	*n*		average age		gender
	screening set	validation set	screening set	validation set	
healthy controls	24	26	69.1	69.1	male
BPH[a]	19	19	68.9	68.9	male
PCa[b] (GS[c] 5–6)	20	17	65.3	65.9	male
PCa[b] (GS[c] 7)	28	30	69.7	64.2	male
PCa[b] (GS[c] 8–10)	25	18	68.1	66.1	male
total	116	110	68.2	64.0	male

[a]Benign prostate hypertrophy. [b]Prostate cancer. [c]Gleason score.

Various sophisticated proteomic techniques have been developed over a couple of decades to explore serum/plasma biomarkers. Most studies utilized focused proteomic technologies to reduce the complexity of serum/plasma proteome, targeting glycosylated proteins,[5,6] peptidome,[7,8] degradome,[9] or minor proteins.[10,11] Low-molecular-weight (LMW) proteome profiling methods have been also employed for biomarker discovery experiments to enrich and detect physiologically important polypeptides, such as cytokines, hormones, and antimicrobial peptides. Although previous LMW enrichment methods significantly enforced the detection limit of small polypeptides, it was difficult to guarantee throughput and reproducibility, which were essential for biomarker studies analyzing multiple clinical specimens.[12−15] For instance, size-exclusion chromatography on HPLC shows better reproducibility but lower throughput due to on-by-one injection of samples. Ultrafiltration spin cartridges show higher throughput but less reproducibility.

Therefore in this study we developed a novel LMW proteome-focusing technology that allows rapid, highly reproducible, and easy-to-operate enrichment of <20 kDa subproteome from crude plasma samples. The principle of this method was repeated purification of denatured undigested protein mixture on 96-well reversed phase chromatography plates. In the light of wide versatility, we named the method quick enrichment of small targets for mass spectrometry (QUEST-MS) technology. Here we applied QUEST-MS technology to both discovery phase and validation phase of prostate cancer biomarker development. Throughout the present study, we show a small neurotransmitter neuropeptide-Y (NPY) as a specific prostate cancer biomarker, which had potential to improve PSA test.

MATERIALS AND METHODS

Reagents

Tris(2-carboxyethyl)phosphine (TCEP), iodoacetamide, and ammonium bicarbonate were purchased from Sigma-Aldrich (Saint Louis, MO). Urea was purchased from GE Healthcare (Buckinghamshire, U.K.). Trypsin Gold was supplied by Promega (Madison, WI). Trifluoroacetic acid (TFA) was purchased from Shimadzu Corporation (Kyoto, Japan).

Plasma Samples

EDTA-plasma samples for biomarker screening (n = 116, Table 1) were collected in Kochi Medical School Hospital. For biomarker validation step, 40 healthy controls were collected in the Cancer Screening Center, The Cancer Institute Hospital of Japanese Foundation for Cancer Research (JFCR). Plasma samples from 20 benign prostate hyperplasia (BPH) patients and 65 prostate cancer patients were collected for validation study in Kochi Medical School Hospital, Kyoto University Hospital, and Iwate Medical University Hospital (Table 1). Plasma from prostate cancer patients and BPH patients were collected before any treatments. Healthy control samples were collected in conjunction with cancer screening. Plasma fraction was separated and stored in the standard operation procedure at hospitals. In brief, withdrawn blood was immediately mixed with EDTA-2K by converting container 10 times and subsequently centrifuged at 1100g for 10 min. The supernatant was aliquoted and stored at −80 °C until use. In all experiments plasma with a single freeze−thaw cycle were used. The research procedure was fully explained, and written informed consent was obtained from all of the patients above. This study was approved by individual institutional ethical committees: The Ethical Committee of RIKEN (Approval code: Yokohama H20-12 and H22-4), Institutional Review Board of Kochi Medical School, JFCR, Kyoto University, and Iwate Medical University.

QUEST-MS purification

On a 96-well polypropylene plate, 20 μL plasma samples were mixed with 80 μL of 10 M urea in 50 mM ammonium bicarbonate. After reduction with 5 mM TCEP at 37 °C for 15 min and alkylation with 25 mM iodoacetamide at room temperature for 15 min, samples were diluted four times with 50 mM ammonium bicarbonate and loaded onto equilibrated Oasis HLB 96-well μElution Plate (2 mg sorbent per well, 30 μm particle size, Waters Corporation, Milford, MA). Here all procedures on solid-phase extraction plates were performed with the custom-made 96-well syringe robot (Supplementary Figure S1 in the Supporting Information). The Oasis plate was prewetted with 500 μL of 70% acetonitrile and equilibrated with 500 μL of 0.1% TFA in 2% acetonitrile at 250 μL/min. Following sample loading at 100 μL/min, plates were washed twice with 500 μL of 0.1% TFA in 2% acetonitrile at 250 μL/min. Proteins were eluted with 100 μL of 40% acetonitrile at 100 μL/min and subsequently diluted five times with 0.1% TFA prior to second Oasis plate purification. The diluted samples were processed with the same conditions as those described in the first purification. The eluates were lyophilized by vacuum spin dryer, followed by digestion with 50 μL of 8 ng/μL Trypsin Gold (Promega) in 50 mM ammonium bicarbonate at 37 °C for 6 h. Digestive reaction was quenched by the addition of 50 μL 1.2% TFA in 4% acetonitrile.

LC/MS/MS Analysis

Following QUEST-MS purification above, 1 μL of sample was analyzed by LTQ-Orbitrap-Velos mass spectrometer (Thermo Fisher Scientific, Waltham, MA) equipped with Ultimate 3000 nanoflow HPLC (Thermo Fisher Scientific). Using 0.1% formic acid as Solvent A and 0.1% formic acid in acetonitrile as Solvent B, peptides were separated on C18 Chip-column (75 μm × 200 mm, Nikkyo Technos, Tokyo, Japan) by the gradient of Solvent B, 2 to 30% for 95 min and 30 to 95% for 15 min at the flow

dx.doi.org/10.1021/pr400547s | J. Proteome Res. 2013, 12, 4497−4506

rate 250 nL/min. The eluted peptides were ionized with the spray voltage 2000 V, and MS data were acquired in a data-dependent fragmentation method in which the survey scan was acquired between m/z 400 and 1600 at the resolution 60 000 with automatic gain control (AGC) target value of 1.0×10^6 ion counts. The top 20 intense precursor ions in each survey scan were subjected to low-resolution MS/MS acquisitions using normal CID scan mode with AGC target value of 5000 ion counts in the linear ion trap.

Label-Free Quantification on Expressionist RefinerMS

The raw data from LTQ-Orbitrap-Velos mass spectrometer were loaded onto Expressionist RefinerMS module (Genedata AG, Basel, Switzerland), which worked on the in-house server system for the subsequent data processing and label-free quantification analysis. The whole workflow of RefinerMS software is shown in Supplementary Figure S2 in the Supporting Information. After setting the Spectrum Grid at every 10 data points on 2D MS chromatogram planes ($x = m/z$ and y = RT), the first chemical noise subtraction was performed using RT Window = 500 scans and Quantile = 80. Following chromatogram smoothing by moving average estimator for every three RT scans in the second chemical noise subtraction, signals less than 1000 intensity were clipped in intensity thresholding. The third and fourth chemical noise subtractions were applied to data using RT structure removal at the minimum RT length = 8 scans and m/z structure removal at the minimum m/z length = 4 points, respectively. The chromatogram grid was set at every 10 scans on noise-subtracted data, followed by chromatogram RT alignment using parameters: m/z window =10 points, RT window =10 scans, gap penalty =1, RT search interval = 2 min, and alignment scheme = pairwise alignment-based tree. Next, the summed peak detection activity detected the peaks on a temporarily averaged chromatogram with parameters as follows: summation window = 2 min, overlap = 50, minimum peak size = 10 scans, maximum merge distance = 4 data points, gap/peak ratio = 10, method = curvature-based peak detection, peak refinement threshold = 5, and consistency filter threshold = 0.6. Finally summed isotope clustering activity grouped isotopic peaks derived from single molecule into an isotope cluster. Here parameters were used as follows: minimum charge = 1, maximum charge = 8, maximum missing peaks = 0, first allowed gap position = 3, ionization = protonation, RT tolerance = 0.1 min, m/z tolerance = 0.05 Da, and minimum cluster size ratio = 0.5.

Extraction of Biomarkers on Expressionist Analyst

Because the specificity of the new biomarker set should be maximized, 153 057 nonredundant detected peptides were filtered by Absent/Present Search algorithm that extracted peptides exhibiting all-or-nothing feature between two groups (healthy control + BPH vs prostate cancer). Peptides detected in at most 1 case among 43 controls and at least 11 cases among 73 prostate cancer patients were selected to be subjected to further validation experiments.

Protein Identification

The SEQUEST database search was performed on Proteome Discoverer 1.3 software (Thermo Fischer Scientific). The MS/MS data were searched against human protein database SwissProt 2012_09 (20 235 sequences) using search parameters as follows: enzyme name = trypsin, precursor mass tolerance = 3 ppm, fragment mass tolerance = 0.8 Da, dynamic modification = oxidation (M), and static modification = carbamidomethyl (C). We accepted peptide identifications that satisfied the false discovery rate (FDR) <1% by Peptide Validator activity in SEQUEST Decoy Database Search. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the data set identifier PXD000383 and DOI 10.6019/PXD000383.

Multiple Reaction Monitoring for NPY

Plasma samples were processed with QUEST-MS procedure as described above, except for the addition of final concentration 10 fmol/μL BSA tryptic digest as an internal control, prior to analysis by 4000 QTRAP mass spectrometer (AB Sciex, Foster City, CA) combined with Agilent 1200 nanoflow HPLC system (Agilent Technologies, Santa Clara, CA). Peptides were separated on C18 Chip-column (75 μm × 200 mm, Nikkyo Technos) using solvent A (0.1% formic acid) and solvent B (0.1% formic acid in acetonitrile). Two-step linear gradient of solvent B was configured from 2 to 30% for 10 min and from 30 to 95% for 5 min at flow rate of 250 nL/min. The multiple reaction monitoring (MRM) transitions specific to NPY and BSA (KYA technologies, Tokyo, Japan) were simultaneously monitored by the MRM mode in Analyst 1.5 software (AB Sciex, Foster City, CA). The instrumental settings were as follows: ionization spray voltage = 2200 V, curtain gas (N_2) = 12 psi, CAD = 4, declustering potential = 70 V, entrance potential = 10 V, Q1 resolution = HIGH, Q3 resolution = LOW, and pause in between = 2 ms. The acquired MRM chromatograms were analyzed with MultiQuant 2.02 software (AB Sciex, Foster City, CA). The mass chromatogram of each transition was smoothed by 1 pt window and quantified by peak area. After optimizing the collision energy (CE) for 12 NPY transitions, quantification was eventually performed using only Q1/Q3 = 466.23/272.20 corresponding to NPY_{81-88} peptide because this transition showed the highest sensitivity. Finally, area of NPY_{81-88} peak was normalized with equally spiked BSA digest as follows: normalized NPY = area (NPY_{81-88}, Q1/Q3 = 466.2/272.2)/area (BSA_{66-75}, Q1/Q3 = 582.3/951.5) × 10 000.

Silver Staining and Image Analysis

To evaluate the efficacy of QUEST-MS purification, we separated one-fifth of purified sample on 16% Tris-tricine SDS-PAGE gel (Life Technologies, Carlsbad, CA). The gel was stained with SilverQuest Silver Staining kit (Life Technologies) by following the manufacturer's instructions. Stained gel was scanned with a GS-800 calibrated densitometer (Bio-Rad Laboratories, Hercules, CA) and analyzed by Image J software.

Immunohistochemical Staining

Immunohistochemical study was carried out using the Ventana automated immunohistochemical systems (Ventana Medical Systems, Tucson, AZ). We used formalin-fixed and paraffin-embedded slice sections of surgical or biopsy specimens from the eight patients with prostate cancer. The eight prostate cancers included three with prostatic intraepithelial neoplasia (PIN) lesions. Sections were incubated with a 1:400 diluted solution of polyclonal anti-NPY antibody (Abcam, Cambridge, U.K.) for 16 min. The automated protocol was based on an indirect biotin–avidin system using a biotinylated universal secondary antibody and diaminobenzidine substrate with hematoxylin counterstaining. The specificity of the binding

was confirmed by negative staining using rabbit nonimmune serum as a primary antibody.

RESULTS

Enrichment of LMW Polypeptides by QUEST-MS Technology

For large-scale biomarker discovery focusing on LMW proteome, we established a rapid and effective enrichment method for <20 kDa proteins in plasma (Figure 1A), termed

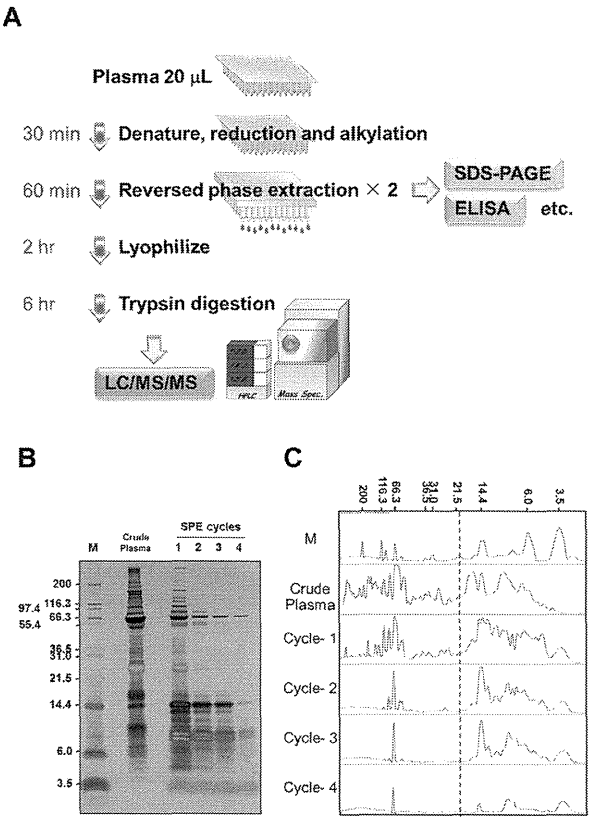


Figure 1. Quick enrichment of small targets for mass spectrometry (QUEST-MS) technology. (A) Schematic workflow of QUEST-MS purification is displayed. After twice so-phase extraction of denatured protein samples within 90 min, low-molecular-weight (LMW) proteins (<20 kDa) are enriched and ready for usual protein assays, such as SDS-PAGE or ELISA. The shotgun or targeted LC/MS/MS analyses can be performed following further tryptic digestion in typically 8 h. (B) Evaluation of LMW protein enrichment by silver-stained 16% Tris-tricine gel. M: molecular weight marker, SPE: solid phase extraction. (C) Result of quantitative densitometry analysis for the gel in panel B is shown. The vertical dashed line indicates 20.0 kDa. More than twice reversed phase purification achieved >90% enrichment of <20 kDa proteins.

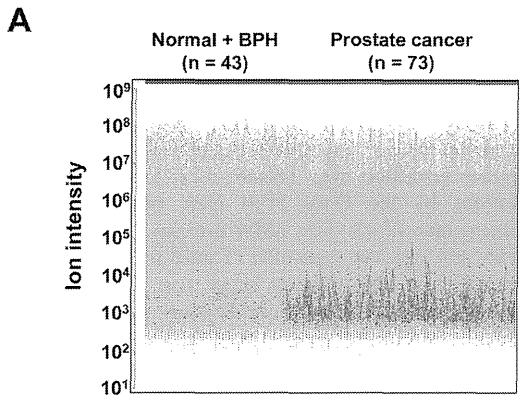
quick enrichment of small targets for mass spectrometry (QUEST-MS). The QUEST-MS procedure included denaturation, reduction, alkylation, and subsequent tandem reversed phase extraction of plasma samples, for which it took only 1.5 h per 96 samples. The dried QUEST-MS-purified samples are now ready for usual protein assays such as SDS-PAGE or ELISA, while they can also be subjected to further mass spectrometric analyses after trypsinization. Crude plasma and purified proteins at each stage were separated on 16% Tris-tricine gel and stained by silver to visualize the efficiency of LMW protein enrichment after the first to fourth reversed-phase extraction by the Oasis HLB plate(Figure 1B). The quantitative density chromatograms of five lanes on the stained gel illustrated effective exclusion of >20 kDa proteins at every solid-phase extraction step (Figure 1C). In particular, two purifications by the Oasis HLB plate resulted in >90% enrichment ratio of <20 kDa LMW proteins. The principle of such easy and clear-cut separation of small proteins was based on pouring away huge proteins larger than pore diameter (8 nm) and preventing elution of middle-range proteins (20–100 kDa) by excess absorption effect to hydrophobic residues on the Oasis polymer beads. Although the third or more repetitions of reversed-phase purification could provide higher purity of LMW proteins, we accepted twice-purified samples for further biomarker discovery and validation experiments from the perspective of retaining advantages in throughput and cost.

LMW Biomarker Screening for Prostate Cancer

To avoid any sampling biases among institutes, we collected 116 plasma samples analyzed in the biomarker screening phase (Table 1) in a single hospital, employing strictly controlled standard operation procedures. The 20 μL of each plasma sample underwent one freeze–thaw step prior to QUEST-MS purification. From LC/MS/MS analysis of 116 purified samples, 153 057 nonredundant peptides were detected and quantified in the Expressionist RefinerMS module, as described in the Materials and Methods section. In particular, this LMW proteome catalog included four PSA-derived peptides $I_{25}VGGWECEK_{33}$ (m/z = 539.25, z = +2), $H_{34}SQPWQVLVASR_{45}$ (m/z = 469.92, z = +3), $A_{48}VCGGVLVHPQWVLTAAHCIR_{68}$ (m/z = 782.08, z = +3), and $L_{126}SEPAELTDAVK_{137}$ (m/z = 636.84, z = +2), which were detectable in 4, 3, 4, and 5 cases from the control group (healthy individuals + BPH patients), whereas they were detected in 15, 13, 10, and 11 cases from the prostate cancer group, respectively. This result suggested that general comparative tests based on the average of samples, such as t test or ANOVA, are not appropriate for the extraction of specific biomarkers from our data set because the valid values of low concentration tumor markers in the control group should be quite few. Therefore to maximize the specificity of new biomarker set, we employed the absent-present search algorithm, which extracted peptides exhibiting all-or-nothing feature between two groups (healthy control + BPH vs prostate cancer, Figure 2A). This search resulted in identification of 189 peaks demonstrating no or one-case detection among 43 controls, while at least 11-case detection among 73 prostate cancer patients (Figure 2B). The extracted peaks above could be considered as candidate biomarkers, which would provide lower false positive rate in addition to similar or higher sensitivity for the diagnosis of prostate cancer patients compared with PSA.

We identified 1126 nonredundant plasma proteins (Supplementary Table S1 in the Supporting Information) from 116 screening samples using Sequest database search analysis (FDR < 0.01). The molecular-weight distribution of 1126 proteins is displayed in Supplementary Figure S3 in the Supporting Information. For qualitative assessment of purified product, 201 out of 1126 identified proteins were <20 kDa. Because quantitative evaluation by SDS-PAGE (Figure 1C) showed >90% enrichment of <20 kDa proteins, most of the 318

dx.doi.org/10.1021/pr400547s | J. Proteome Res. 2013, 12, 4497–4506



B

		Prostate cancer								
		19	18	17	16	15	14	13	12	11
Normal + BPH	0	0	0	0	0	1	3	2	11	19
	1	1	2	2	3	8	10	21	44	62

Figure 2. Prostate cancer biomarker screening by the absent/present search method on Expressionist proteome server. (A) 189 peptides out of 153 057 total peptides showed prostate cancer-specific detection patterns, demonstrating no or only one-case detection in the control group. The quantitative information of 189 peptides is indicated by red dots or lines on the 2D plane composed of 110 individuals as x axis and the peak intensity in LC/MS/MS as y axis. Normal: 24 healthy controls, BPH: 19 benign prostate hypertrophy patients. (B) Classification of 189 candidate biomarker peptides based on the frequency of detection. For instance, 62 peptides were detectable in 11 prostate cancer samples, while detected in 1 control sample (normal or BPH).

in cerebral nervous system and prostate gland,[19] we subjected neuropeptide Y (NPY) to the next validation experiments.

NPY Expression in Prostate Cancer Tissues

NPY immunostaining patterns of prostate cancer frozen tissues at Gleason score 3 + 4 (Figure 3A,D) or 4 + 5 (Figure 3B,E)

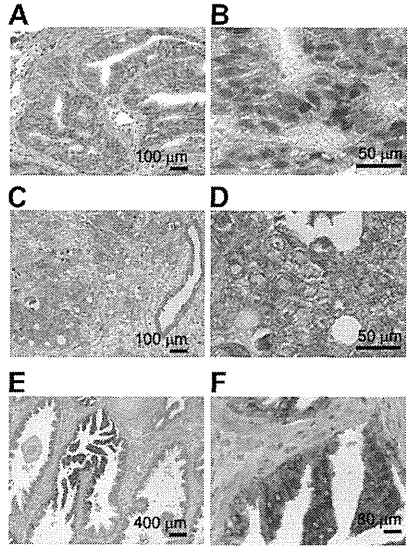


Figure 3. Immunohistochemical staining analysis for pro-neuropeptide Y (NPY). The NPY immunostaining patterns of prostate cancer at Gleason Score 3 + 4 (A) and 4 + 5 (C), respectively. Strong positive immunostaining was observed in the cytoplasm of prostate cancer cells, whereas a weak immunopositivity was found in noncancerous prostate epithelium. (E) Cytoplasmic immunoreactivity with anti-NPY antibody of the high-grade PIN precursor cells. (B,D,F) Five-fold magnified view of A, C, and E, respectively. Scale bars are shown at the right bottom of panels.

proteins more than 100 kDa (as UniProt database values) might exist in cleaved or degraded forms in purified product. Here, in addition to the detection of PSA, we observed low abundant chemokines (e.g., CCL14, CCL18, CXCL4, and CXCL7), hormones (e.g., gastric inhibitory polypeptide and inhibin β, whose concentrations were reported as 200–300 pg/mL[16] and 10–200 pg/mL,[17] respectively), hypotensive peptide adrenomedullin (50–100 pg/mL),[18] and others. By matching 189 candidate biomarker peptides against the imported protein identification data set on the Expressionist proteome server system, three peptides listed in Table 2 were successfully identified. From the aspect of highly specific mRNA expression are shown. Immunoreactivity with anti-NPY antibody was observed in prostate cancer tissues, exhibiting strong positive immunostaining in the cytoplasm of prostate cancer cells, whereas a weak immunopositivity was found in noncancerous prostate epithelium. Figure 3C,F further indicated that the cytoplasmic immunoreactivity with anti-NPY antibody could be also observed in PIN precursor cells. These results suggested that NPY expression was significantly upregulated at neoplastic lesions in prostate tissue. The staining intensity of NPY seemed to be varied according to Gleason grade of tumor cells within a field.

Table 2. List of Identified Biomarker Candidates for Prostate Cancer

				detection frequency in absent/present search	
UniProt accession number	UniProt ID	protein name	sequence[a]	control[b]	PCa[c]
P01303	NPY_HUMAN	pro-neuropeptide Y	S^{68}SPETLISDLLMR^{80}	0	11
P02751	FINC_HUMAN	fibronectin	Q^{1041}YNVGPSVSK^{1050}	1	11
Q16134	ETFD_HUMAN	electron transfer flavoprotein-ubiquinone oxidoreductase	G^{213}IATNDVGIQK^{223}	1	11

[a]Numbers in the sequences indicate the amino acid numbers. [b]Number of cases detected among healthy controls or BPH patients. [c]Number of cases detected among prostate cancer patients.

dx.doi.org/10.1021/pr400547s | J. Proteome Res. 2013, 12, 4497–4506

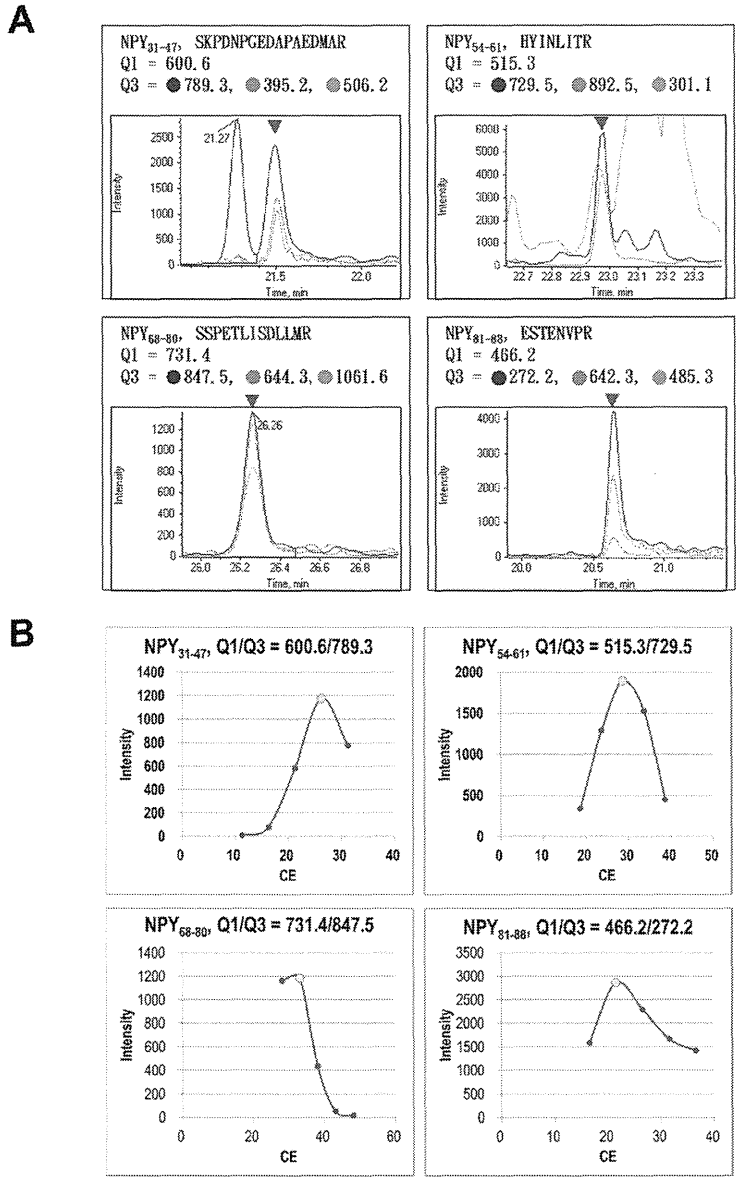


Figure 4. Optimization of MRM conditions for NPY used in the biomarker validation experiment. (A) Selection of the appropriate MRM transition for further quantification analysis. Three transitions were tested for each of four NPY-derived peptides identified in the biomarker screening step. The numbers on the right side of Q1 or Q3 indicate the target setting of m/z in the quadrupole 1 or 3, respectively. (B) Transition showing the highest quality (S/N) of MRM chromatogram was measured with five distinct conditions of collision energy (CE, eV). The CE setting providing the highest intensity of chromatogram peak is indicated by yellow circle.

MRM-Based Replication Analysis for NPY

To evaluate the variability associated with different hospitals or analytical methods, we conducted further validation experiment for NPY using an independent sample set and another mass spectrometric quantification method MRM. Here we replaced healthy controls from Kochi University hospital with ones from JFCR and added prostate disease samples from Kyoto University and Iwate Medical University (Table 1). The 110 QUEST-MS-purified samples were then analyzed in a 4000 QTRAP mass spectrometer by targeting four distinct peptides derived from NPY. Prior to MRM measurement of 110 validation samples, we first searched the optimum transition and CE out of three transitions for each four NPY peptide, which could provide the highest intensity of MRM chromatogram peaks (Figure 4A,B). From the analysis of case-1 in Supplementary Figure S4 in the Supporting Information, the transition Q1/Q3 = 466.2/272.2 corresponding to NPY_{81-88} peptide at CE = 17 eV was found to be the most sensitive reporter ion pair for NPY. Because only NPY_{81-88} transition had enough sensitivity to detect NPY with even normal and BPH control levels, the following NPY quantification analysis was performed by this transition. Additionally, 10 fmol of digested BSA was spiked to each sample and used as the internal control to normalize the interanalytical variability. The quantitative reproducibility of MRM-based NPY measurement

dx.doi.org/10.1021/pr400547s | J. Proteome Res. 2013, 12, 4497−4506

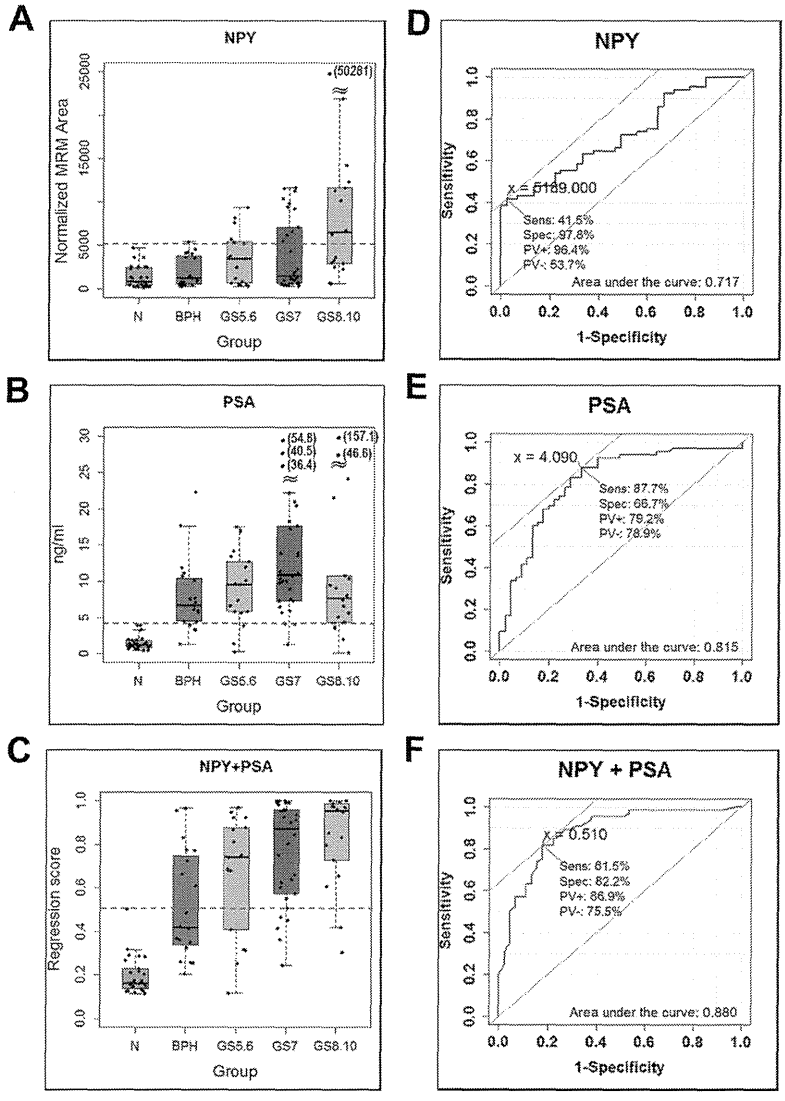


Figure 5. Results of MRM-based validation experiment for NPY using 110 plasma samples. The quantified concentration of NPY by MRM or PSA by ELISA is shown by the box plot (A) or (B), respectively. The combination biomarker of NPY and PSA was constructed by the modified logistic regression method and utilized for the classification of five groups (C). The red dashed lines indicate the thresholds between control group (N + BPH) and prostate cancer group, calculated by the ROC curve analysis in panels D−F. The outliers are displayed separately in the upper area of plots, accompanying each concentration. N: healthy controls, BPH: benign prostate hypertrophy, GS: Gleason score. ROC curves were constructed with plasma concentration of NPY (D), PSA (E), or the regression score of combination marker NPY + PSA (F) by use of 110 validation plasma samples. The point of tangency indicated by "x =" shows the optimum threshold value provided by each biomarker. Sens.: sensitivity, Spec.: specificity, PV+: positive predictive values, PV−: negative predictive values.

was then evaluated by analyzing three plasma samples at six distinct time points in triplicated manner. Supplementary Figure S4 in the Supporting Information showed that observed coefficient of variation (CV) was consistently smaller than 3.5%. Then, by use of optimized MRM conditions above, plasma NPY was measured by relative quantification from 110 cases and displayed on a box plot in Figure 5A, demonstrating much higher diagnostic specificity between BPH patients and Gleason score 5 to 6 prostate cancer patients (p = 0.0396, t test) in comparison with PSA test (p = 0.2746, t test) (Figure 5B). This observation successfully confirmed the prostate-cancer-specific detection of NPY in the absent/present search analysis at the biomarker screening phase. Despite the high specificity, because the sensitivity of NPY appeared not to be sufficient enough for a singly used biomarker, we intended to establish an integrative diagnostic model of NPY and PSA by use of modified logistic regression method (Figure 5C). The logistic regression scores were calculated so that they could provide the maximum area under the curve (AUC) of ROC curve as previously described.[20] Figure 5D−F showed the comparison of three ROC curves generated with NPY (area of MRM chromatogram), PSA (values measured by ELISA), or NPY + PSA (logistic regression scores) of 110 plasma samples, respectively. Whereas the sensitivity and specificity of PSA in distinguishing 65 prostate cancer patients from 45 controls (25 healthy controls and 20 BPH patients) were 87.7 and 66.7%,

dx.doi.org/10.1021/pr400547s | J. Proteome Res. 2013, 12, 4497−4506

respectively, those of the NPY-PSA combination biomarker reached 81.5 and 82.2%, respectively. This result clearly suggested that NPY improved the poor specificity of PSA test without significant loss of its sensitivity, which would lead to reduction of overdiagnosis and unnecessary biopsies, eventually leading to alleviation of both physical and mental stress for men.

DISCUSSION

For the purpose of identifying blood biomarkers for early diagnosis of cancers, comprehensive detection of submicrogram/milliliter proteins should be fundamental because the amount of potential tumor marker proteins secreted from small malignant tissues would be very subtle, even allowing for relatively rapid accumulation of particular proteins which show long half-life in blood, such as CEA (7−14 days),[21] AFP (6.2 days),[22] or CA125 (6 days).[23] To achieve such high sensitivity in plasma proteomics, we focused on <20 kDa LMW subproteome of plasma because most of interfering abundant proteins were found in larger molecular weight area of plasma proteome. The QUEST-MS technology we developed in the present study enabled us not only to enrich the <20 kDa fractions rapidly but also to identify a number of plasma minor components. These facts strongly assured that the sensitivity provided by QUEST-MS technology was high enough to discover very low concentration of biomarker proteins derived from early diseases utilizing any mass spectrometric approaches. This technology can also enhance the detection level of various protein assays, such as ELISA, Western blotting, or HPLC-based analysis targeting small proteins. The considerable benefits of QUEST-MS technology in the clinical application were reproducibility and quickness of experimental procedures. Unlike size exclusion chromatography or ultrafiltration methods, all processes can be executed rapidly by automatic pipetting robots equipped with 96-well syringe pumps (Supplementary Figure S1 in the Supporting Information).

In this report, we applied QUEST-MS technology to the plasma biomarker discovery for prostate cancer by analyzing 116 samples. The whole analytical steps were accomplished within a couple of weeks, involving 1 day of QUEST-MS purification and 12 days of individual LC/MS/MS analyses. The subsequent absent/present search was successful in extracting 189 candidate biomarker peptides showing prostate cancer-specific detection patterns; however, only three of them were sequenced by Sequest database search analysis despite identification of 1126 proteins in total. This poor identification rate might be attributed to quite low concentrations of 189 candidate peaks. Actually, the S/N ratio of mass spectrum peaks for these peptides was only 3−10, resulting in missing MS/MS acquisitions or insufficient quality of MS/MS spectra for database search. As described in the previous paragraph, it is no wonder that early diagnosis biomarkers would be found in the lowest intensity area of mass spectrometric data and the sequence identifications were not always straightforward. Therefore, MudPIT (multidimensional protein identification technology) approaches employing more intensive prefractionation of analytes should be necessary to enforce the identification efficacy, even though the detection in the survey scans has already been achieved.

The new prostate cancer biomarker NPY on which we focused in the present study is one of the well-known neurotransmitter polypeptide composed of 36 amino acids as the activated from in blood. The prepro-NPY (97 amino acids) is synthesized from NPY gene and secreted into extracellular region after removal of the signal peptide NPY_{1-28} (pro-NPY). In the next stage, the C-flanking peptide of NPY (CPON, NPY_{68-97}) is cleaved from pro-NPY by proconverting enzymes, resulting in NPY_{29-67}. Following further cleavage at the C-terminus of Gly^{65} by carboxypeptidase-like enzyme, the Gly^{65} was finally truncated by peptidylglycine-amidating monooxygenase to generate mature NPY_{29-64}. In our biomarker screening data, only NPY_{68-80} tryptic digest was identified, which demonstrated the detection profile as follows (healthy controls + BPH)/(prostate cancer patients) = 0/11. However the MRM-based validation of four distinct NPY tryptic digests, including NPY_{31-47}, NPY_{54-61}, NPY_{68-80}, and NPY_{81-88}, denoted the same tendency of plasma concentration among 110 individuals. These facts could provide two important insights concerning the limit of detection and the dynamic feature of NPY in prostate cancer patients. At first, all of NPY_{31-47}, NPY_{54-61}, and NPY_{81-88} peptides were not involved in 189 biomarker candidates because they were masked by coeluted other peptides and detected only infrequently. Second, the isoform upregulated in prostate cancer patients' plasma was the inactive form pro-NPY (NPY_{29-97}) whose C-terminus had not been removed.

Although major physiological functions of active NPY_{29-64} are known to be strong appetite stimulant,[24] vasoconstrictive effect,[25] and regulation of stress behaviors,[26] the association between NPY level and cancer progression remains controversial. At least the NPY seems to be involved in the development of specific tumors, including neural crest-derived tumors and breast and prostate cancers by facilitating proliferation, invasion, metastasis, and angiogenesis.[27] Importantly, it was reported that overexpression of NPY in prostate cancer tissues had significant correlation with poor prognosis of patients.[28] These reports could support the evidence that a particular group of prostate cancer patients exhibit an aberrantly high level of plasma NPY even at very early grade of cancer such as Gleason score 5 or 6. In fact, the addition of NPY values to PSA test using logistic regression model significantly improved the diagnostic specificity between BPH patients and Gleason score 5 to6 prostate cancer patients (Figure 5).

Recently, two innovative gene diagnostics for prostate cancer, urine PCA3 and TMPRSS2:ERG, have been under clinical reviews. Large-scale trials of urine PCA3 test in the United States (n = 466) and Japan (n = 633) yielded 77.5% sensitivity and 57.1% specificity (PCA3 score threshold = 25)[29] or 66.5% sensitivity and 71.6% specificity (PCA3 score threshold = 35),[30] respectively. TMPRSS2:ERG fusion gene was found in 41 or 43% of primary prostate cancer tissues (n = 59) or castration-resistant prostate cancer tissues (n = 82), respectively.[31] A multivariate regression analysis combining expression levels of PCA3, GOLPH2, SPINK1, and TMPRSS2:ERG in urine cells outperformed serum PSA, resulting in 65.9% sensitivity and 76.0% specificity (n = 234).[32] Even in comparison with these gene diagnostics, our combinational biomarker NPY+PSA illustrated better performance to distinguish prostate cancer from noncancer population (81.5% sensitivity and 82.2% specificity).

CONCLUSIONS

In summary, we described the development of versatile LMW-focusing proteomics technology QUEST-MS and the application of this method to plasma biomarker discovery for prostate

cancer. QUEST-MS technology provides an adequate reduction in sample complexity to improve detection of low abundance proteins. Although here we utilized Oasis HLB μElution plates for QUEST-MS purification, other reversed-phase chromatographic materials with different functional moieties or pore sizes would expand the variety of purified molecules. In future applications, the automated QUEST-MS procedure will drastically facilitate sensitive and high-throughput analysis of small bioactive polypeptides, such as hormones, neurotransmitters, or cytokines, from any body fluids including serum, cerebrospinal fluid, or urine.

■ ASSOCIATED CONTENT

Supporting Information

Automatic 96-well syringe robot used for QUEST-MS purification, workflow for the label-free quantification analysis on the Expressionist RefinerMS module, molecular weight distribution of identified proteins from QUEST-MS-purified plasma samples, evaluation of variability for MRM-based plasma NPY measurements, and a list of 1126 nonredundant proteins identified from biomarker screening phase. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +81-3-6409-2306. Fax: +81-3-5449-5785. E-mail: k-ueda@riken.jp.

Notes

The authors declare the following competing financial interest(s): A.T. and T.A.S. are employees of Shimadzu Corporation. M.N. is employee of Toppan Printing Co., Ltd.

■ ACKNOWLEDGMENTS

This work was supported by funding to Koji Ueda. from Grant-in-Aid for Young Scientists (A), The Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. Shimadzu Corporation and Toppan Printing Co., Ltd. also financially supported the present study.

■ ABBREVIATIONS

QUEST-MS, quick enrichment of small targets for mass spectrometry; BPH, benign prostate hypertrophy; PSA, prostate-specific antigen; NPY, neuropeptide-Y; PIN, prostatic intraepithelial neoplasia; MRM, multiple reaction monitoring; LMW, low-molecular-weight

■ REFERENCES

(1) Ablin, R. J.; Pfeiffer, L.; Gonder, M. J.; Soanes, W. A. Precipitating antibody in the sera of patients treated cryosurgically for carcinoma of the prostate. *Exp. Med. Surg.* **1969**, *27* (4), 406–10.

(2) Catalona, W. J.; Smith, D. S.; Ratliff, T. L.; Dodds, K. M.; Coplen, D. E.; Yuan, J. J.; Petros, J. A.; Andriole, G. L. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N. Engl. J. Med.* **1991**, *324* (17), 1156–61.

(3) Chou, R.; Croswell, J. M.; Dana, T.; Bougatsos, C.; Blazina, I.; Fu, R.; Gleitsmann, K.; Koenig, H. C.; Lam, C.; Maltz, A.; Rugge, J. B.; Lin, K. Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. *Ann. Intern. Med.* **2011**, *155* (11), 762–71.

(4) Brett, A. S.; Ablin, R. J. Prostate-cancer screening–what the U.S. Preventive Services Task Force left out. *N. Engl. J. Med.* **2011**, *365* (21), 1949–51.

(5) Kurogochi, M.; Matsushista, T.; Amano, M.; Furukawa, J.; Shinohara, Y.; Aoshima, M.; Nishimura, S. Sialic acid-focused quantitative mouse serum glycoproteomics by multiple reaction monitoring assay. *Mol. Cell. Proteomics* **2010**, *9* (11), 2354–68.

(6) Ueda, K.; Takami, S.; Saichi, N.; Daigo, Y.; Ishikawa, N.; Kohno, N.; Katsumata, M.; Yamane, A.; Ota, M.; Sato, T. A.; Nakamura, Y.; Nakagawa, H. Development of serum glycoproteomic profiling technique; simultaneous identification of glycosylation sites and site-specific quantification of glycan structure changes. *Mol. Cell. Proteomics* **2010**, *9* (9), 1819–28.

(7) Villanueva, J.; Martorella, A. J.; Lawlor, K.; Philip, J.; Fleisher, M.; Robbins, R. J.; Tempst, P. Serum peptidome patterns that distinguish metastatic thyroid carcinoma from cancer-free controls are unbiased by gender and age. *Mol. Cell. Proteomics* **2006**, *5* (10), 1840–52.

(8) Ueda, K.; Saichi, N.; Takami, S.; Kang, D.; Toyama, A.; Daigo, Y.; Ishikawa, N.; Kohno, N.; Tamura, K.; Shuin, T.; Nakayama, M.; Sato, T. A.; Nakamura, Y.; Nakagawa, H. A comprehensive peptidome profiling technology for the identification of early detection biomarkers for lung adenocarcinoma. *PloS One* **2011**, *6* (4), e18567.

(9) Shen, Y.; Liu, T.; Tolic, N.; Petritis, B. O.; Zhao, R.; Moore, R. J.; Purvine, S. O.; Camp, D. G.; Smith, R. D. Strategy for degradomic-peptidomic analysis of human blood plasma. *J. Proteome Res.* **2010**, *9* (5), 2339–46.

(10) Liu, T.; Qian, W. J.; Mottaz, H. M.; Gritsenko, M. A.; Norbeck, A. D.; Moore, R. J.; Purvine, S. O.; Camp, D. G., 2nd; Smith, R. D. Evaluation of multiprotein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. *Mol. Cell. Proteomics* **2006**, *5* (11), 2167–74.

(11) Mouton-Barbosa, E.; Roux-Dalvai, F.; Bouyssie, D.; Berger, F.; Schmidt, E.; Righetti, P. G.; Guerrier, L.; Boschetti, E.; Burlet-Schiltz, O.; Monsarrat, B.; Gonzalez de Peredo, A. In-depth exploration of cerebrospinal fluid by combining peptide ligand library treatment and label-free protein quantification. *Mol. Cell. Proteomics* **2010**, *9* (5), 1006–21.

(12) Capriotti, A. L.; Caruso, G.; Cavaliere, C.; Piovesana, S.; Samperi, R.; Lagana, A. Comparison of three different enrichment strategies for serum low molecular weight protein identification using shotgun proteomics approach. *Anal. Chim. Acta* **2012**, *740*, 58–65.

(13) Hood, B. L.; Lucas, D. A.; Kim, G.; Chan, K. C.; Blonder, J.; Issaq, H. J.; Veenstra, T. D.; Conrads, T. P.; Pollet, I.; Karsan, A. Quantitative analysis of the low molecular weight serum proteome using 18O stable isotope labeling in a lung tumor xenograft mouse model. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (8), 1221–30.

(14) Lassout, O.; Pastor, C. M.; Fetaud-Lapierre, V.; Hochstrasser, D. F.; Frossard, J. L.; Lescuyer, P. Analysis of the pancreatic low molecular weight proteome in an animal model of acute pancreatitis. *J. Proteome Res.* **2010**, *9* (9), 4535–44.

(15) Wu, J.; An, Y.; Pu, H.; Shan, Y.; Ren, X.; An, M.; Wang, Q.; Wei, S.; Ji, J. Enrichment of serum low-molecular-weight proteins using C18 absorbent under urea/dithiothreitol denatured environment. *Anal. Biochem.* **2010**, *398* (1), 34–44.

(16) Wang, Y. Y.; Lee, C. T.; Lu, C. L.; Chen, C. Y.; Chang, F. Y.; Lee, S. D.; Doong, M. L.; Wang, P. S. Gastric inhibitory polypeptide appears less important in mediating acid secretion. *Hepato-Gastroenterology* **1999**, *46* (27), 2105–9.

(17) Sehested, A.; Juul, A. A.; Andersson, A. M.; Petersen, J. H.; Jensen, T. K.; Muller, J.; Skakkebaek, N. E. Serum inhibin A and inhibin B in healthy prepubertal, pubertal, and adolescent girls and adult women: relation to age, stage of puberty, menstrual cycle, follicle-stimulating hormone, luteinizing hormone, and estradiol levels. *J. Clin. Endocrinol. Metab.* **2000**, *85* (4), 1634–40.

(18) Di Iorio, R.; Marinoni, E.; Letizia, C.; Villaccio, B.; Alberini, A.; Cosmi, E. V. Adrenomedullin production is increased in normal human pregnancy. *Eur. J. Endocrinol.* **1999**, *140* (3), 201–6.

(19) Su, A. I.; Wiltshire, T.; Batalov, S.; Lapp, H.; Ching, K. A.; Block, D.; Zhang, J.; Soden, R.; Hayakawa, M.; Kreiman, G.; Cooke, M. P.; Walker, J. R.; Hogenesch, J. B. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (16), 6062–7.

dx.doi.org/10.1021/pr400547s | J. Proteome Res. 2013, 12, 4497–4506

(20) Pepe, M. S.; Cai, T.; Longton, G. Combining predictors for classification using the area under the receiver operating characteristic curve. *Biometrics* **2006**, *62* (1), 221–9.
(21) Choi, J. S.; Min, J. S. Significance of postoperative serum level of carcinoembryonic antigen (CEA) and actual half life of CEA in colorectal cancer patients. *Yonsei Med. J.* **1997**, *38* (1), 1–7.
(22) Gerl, A.; Lamerz, R.; Clemm, C.; Mann, K.; Hartenstein, R.; Wilmanns, W. Does serum tumor marker half-life complement pretreatment risk stratification in metastatic nonseminomatous germ cell tumors? *Clin. Cancer Res.* **1996**, *2* (9), 1565–70.
(23) Meyer, T.; Rustin, G. J. Role of tumour markers in monitoring epithelial ovarian cancer. *Br. J. Cancer* **2000**, *82* (9), 1535–8.
(24) Zhang, L.; Bijker, M. S.; Herzog, H. The neuropeptide Y system: pathophysiological and therapeutic implications in obesity and cancer. *Pharmacol. Ther.* **2011**, *131* (1), 91–113.
(25) Hirsch, D.; Zukowska, Z. NPY and stress 30 years later: the peripheral view. *Cell. Mol. Neurobiol.* **2012**, *32* (5), 645–59.
(26) Sah, R.; Geracioti, T. D. Neuropeptide Y and posttraumatic stress disorder. *Mol. Psychiatry* **2013**, *18* (6), 646–55.
(27) Ruscica, M.; Dozio, E.; Motta, M.; Magni, P. Relevance of the neuropeptide Y system in the biology of cancer progression. *Curr. Top. Med. Chem.* **2007**, *7* (17), 1682–91.
(28) Rasiah, K. K.; Kench, J. G.; Gardiner-Garden, M.; Biankin, A. V.; Golovsky, D.; Brenner, P. C.; Kooner, R.; O'Neill, G, F.; Turner, J. J.; Delprado, W.; Lee, C. S.; Brown, D. A.; Breit, S. N.; Grygiel, J. J.; Horvath, L. G.; Stricker, P. D.; Sutherland, R. L.; Henshall, S. M. Aberrant neuropeptide Y and macrophage inhibitory cytokine-1 expression are early events in prostate cancer development and are associated with poor prognosis. *Cancer Epidemiol., Biomarkers Prev.* **2006**, *15* (4), 711–6.
(29) Gittelman, M.; Hertzman, B.; Bailen, J.; Williams, T.; Koziol, I.; Henderson, R. J.; Efros, M.; Bidair, M.; Ward, J. F. PCA3 molecular urine test as a predictor of repeat prostate biopsy outcome in men with previous negative biopsies: A prospective multicenter clinical study. *J. Urol.* **2013**, *190* (1), 64–69.
(30) Ochiai, A.; Okihara, K.; Kamoi, K.; Oikawa, T.; Shimazui, T.; Murayama, S. I.; Tomita, K.; Umekawa, T.; Uemura, H.; Miki, T. Clinical utility of the prostate cancer gene 3 (PCA3) urine assay in Japanese men undergoing prostate biopsy. *BJU Int.* **2013**, *111* (6), 928–33.
(31) Qu, X.; Randhawa, G.; Friedman, C.; O'Hara-Larrivee, S.; Kroeger, K.; Dumpit, R.; True, L.; Vakar-Lopez, F.; Porter, C.; Vessella, R.; Nelson, P.; Fang, M. A novel four-color fluorescence in situ hybridization assay for the detection of TMPRSS2 and ERG rearrangements in prostate cancer. *Cancer Genet.* **2013**, *206* (1–2), 1–11.
(32) Laxman, B.; Morris, D. S.; Yu, J.; Siddiqui, J.; Cao, J.; Mehra, R.; Lonigro, R. J.; Tsodikov, A.; Wei, J. T.; Tomlins, S. A.; Chinnaiyan, A. M. A first-generation multiplex biomarker analysis of urine for the early detection of prostate cancer. *Cancer Res.* **2008**, *68* (3), 645–9.

dx.doi.org/10.1021/pr400547s | *J. Proteome Res.* 2013, 12, 4497–4506

REVIEW

Glycoproteomic strategies: From discovery to clinical application of cancer carbohydrate biomarkers

Koji Ueda

Laboratory for Biomarker Development, Center for Genomic Medicine, RIKEN, Minato-ku, Tokyo, Japan

Received: November 29, 2012
Accepted: December 27, 2012

Carbohydrate antigens are the most frequently and traditionally used biomarkers for cancer, such as CA19–9, CA125, DUPAN-II, AFP-L3, and many others. The diagnostic potential of them was simply based on the cancer-specific alterations of glycan structures on particular glycoproteins in serum/plasma. In spite of the facts that glycosylation disorders are feasible for cancer biomarkers and glycomic analysis technologies to explore them have been rapidly developed, it remains difficult to sensitively screen glycan structure changes on cancer-associated glycoproteins from clinical specimens. Moreover, a lot of additional issues should be appropriately addressed for the clinical application of newly identified glycosylation biomarkers, including analytical throughput, quantitative confirmation of structural changes, and biological explanation for the alterations. In the last decade, significant improvement of mass spectrometric techniques is being made in the aspects of both hardware spec and preanalytical purification procedures for glycoprotein analysis. Here we review potential approaches to perform comprehensive analysis of glycoproteomic biomarker screening from serum/plasma and to realize high-throughput validation of site-specific oligosaccharide variations. The power and problems of mass spectrometric applications on the clinical use of carbohydrate biomarkers are also discussed in this review.

Keywords:
Biomarker / Cancer / Glycoproteomics / Mass spectrometry / Serum

1 Introduction

In order to survey tumor-derived biomarkers in serum, not only glycoproteomics but all other focusing technologies are essential, which effectively reduce the complexity of samples by removing unnecessary major serum proteins [1, 2]. Such focused proteomics approaches would include peptidomics [3–6], degradomics [7–11], immunodepletion of abundant proteins [12–16], ProteoMiner purification [17–22], and rare amino acid capturing (e.g. iCAT tag [23–28] for cysteine or 2-nitrobenzenesulfenyl tag for tryptophan [29–34]). Since concentration of biomarker molecules released from early-stage cancer tissues (generally less than 20 mm diameter) are extremely low, appropriate focused proteomics technologies can greatly help to detect such sub-μg/mL proteins in sera. In particular, glycosylation-focused proteomics could contribute to both improving analytical depth and interpreting roles of carcinogenesis-associated glycosylations, which might lead to further understanding molecular mechanisms of malignant transformation or metastasis. More importantly, particular alterations of glycan structures on secreted glycoproteins specifically reflect the site of original organs or cell types. This aspect of glycosylation biomarkers enforces the specificity of diagnosis in the cases distinguishing cancer from benign diseases (e.g. lung cancer from chronic obstructive pulmonary disease) or defining origin of cancer. Despite the benefits described above, there have been lots of technical difficulties in the discovery of cancer-associated glycosylation microheterogeneities from clinical specimens as well as clinical application of those. Indeed, the latest series of mass spectrometers can provide much deeper and wider knowledge of proteins compared to those a decade ago. However, it has been realized that mass spectrometric analysis of lectin-purified serum proteins, which is one of the simplest ways of glycoproteomics, is far from sufficient to

Correspondence: Dr. Koji Ueda, Laboratory for Biomarker Development, Center for Genomic Medicine, RIKEN, Institute of Medical Science, General Research Building 6F, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
E-mail: k-ueda@riken.jp
Fax: +81-3-5449-5785

Abbreviations: Erexim, energy-resolved oxonium ion monitoring; **GalNAz**, tetra-aceylated-*N*-azidoacetylgalactosamine; **GpIb**, glycoprotein Ib; **IGEL**, isotopic glycosidase elution and labeling on lectin-column chromatography; **WGA**, wheat germ agglutinin

identify glycosylation biomarkers for early cancer diagnosis. Here we discuss about issues and resolutions in the use of lectins by introducing methods for sample preparation prior to lectin chromatography. In addition to lectin-based approaches, which have been most frequently utilized in glycomics, other sophisticated glycocapturing technologies are reviewed. Because the concept required for biomarker discovery phase or preclinical validation phase is fundamentally different, the adequate methodologies for each are separately described.

2 Glycoproteomics for biomarker screening

2.1 Lectins or chemicals

Nowadays, state-of-the-art glycoproteomic technologies have been developed to enrich glycoproteins or glycopeptides from crude serum samples. These technologies are separated into lectin-based methods and chemical-based methods, in principle. Which is better for the purpose of carbohydrate-targeting tumor marker discovery? If we intend to identify glycan structure changes as tumor markers, rather than concentration of core proteins, lectin is the only enrichment tool recognizing specific oligosaccharide linkages, excepting sialic acid specific chemistry reverse glycoblotting [35, 36]. The chemical enrichment of glycopeptides, such as using hydrazide chemistry [37–42], boronic acid [43–45], or hydrophilic interactions [46–48], certainly exhibits rigid interaction with glycan moieties, whereas most of lectin–glycan interactions are fragile [49]. However, all of the chemical approaches above are based on covalent or affinity bond with rich hydroxyl groups on oligosaccharides, resulting in comprehensive and structure-unspecific capture of glycopeptides. Therefore, the chemical route is inadvisable for glycan structure-targeting biomarker discovery. From the view of such features, we would like to focus on lectin-based glycopeptide enrichment methods in the following sections.

2.2 Using lectin column chromatography for glycoproteomics

Assuming that the eluate of lectin column chromatography would be analyzed in LC/MS/MS, it must be a critical issue whether we load proteins or digested peptides to lectin columns. When undigested serum proteins are purified with lectin column chromatography and eluted by hapten sugars, targeted glycoproteins would be eluted with a lot of nonspecific proteins and high concentration of hapten sugars, which could not be appropriate for mass spectrometric analysis (Fig. 1A). The large amount of nonglycosylated protein elution is mainly caused by the limitation of solvent used in lectin column chromatography, with which the use of detergents, high salts, and organic solvents are not compatible. Furthermore, since most of serum proteins form complexes, coelution of intact binding proteins is inevitable. More importantly, if enriched glycoproteins would have multiple glycosylation sites, it is hardly distinguishable which glycosylation sites might be associated with state of cancer.

On the other hand, when digested peptides are loaded to lectin columns, the purification efficiency of glycopeptides is relatively high due to the elimination of protein–protein interaction effect in the samples (Fig. 1B). However, elution of glycopeptides using hapten sugars still results in significant contamination of 200–1000 mM sugars in the eluate, which is not adequate for direct injection to LC/MS/MS. In addition to this fact, mass spectrometric analysis of eluted glycopeptides is also inefficient for comprehensive studies because automated protein identification by database search is impossible for glycopeptides, for which additional deglycosylation steps would be finally required.

These characteristics of lectins on glycoproteomics have led to recent development of glycopeptidase elution technologies [46, 50]. After binding glycopeptides on lectin columns, PNGase-F elution in volatile salt buffers allows highly specific elution of only originally glycosylated peptides. After lyophilizing eluate, the final product consists of completely deglycosylated peptides with no contaminant salts, which could be directly injected into LC/MS instruments and subjected to usual database search analysis on Mascot or Sequest software. Using this type of enzymatic elution procedure on click maltose HILIC beads, Zhu et al. successfully identified 92, 178, and 221 unique *N*-glycosylation sites from 10 nL, 100 nL, and 1 μL of human serum, respectively [46]. They effectively excluded desalting, buffer exchanging, and lyophilization steps to finish all preanalytical procedures by spin columns within 1.5 h.

Regarding peptide sequencing of enzymatically deglycosylated peptides on database search analysis, *N*-glycosylation sites are recognized as aspartic acid residues converted from asparagine residues by PNGase-F (Fig. 1C). However, chemically identical conversion may artificially occur on asparagine residues known as deamidation of asparagine. To eliminate the false-positive identification of *N*-glycosylation sites by deamidation, we can utilize the PNGase-F reaction in heavy water ($H_2{}^{18}O$) [50–57]. The incorporation of ^{18}O into glycosylated asparagine residues by PNGase-F induces generation of 3 Da increased asparagine residues, providing N-glycosylation site-specific stable isotope tags on peptides (Fig. 1C). Our team recently integrated on-column PNGase-F elution with the ^{18}O stable isotope labeling method and reported as an effective glycoproteomic biomarker screening technology, named isotopic glycosidase elution and labeling on lectin-column chromatography (IGEL) [50].

Hence, issues on biochemical properties of lectins for glycoproteomics can be overcome by employing both protease digestion before lectin column purification and on-column glycopeptidase elution of peptides.

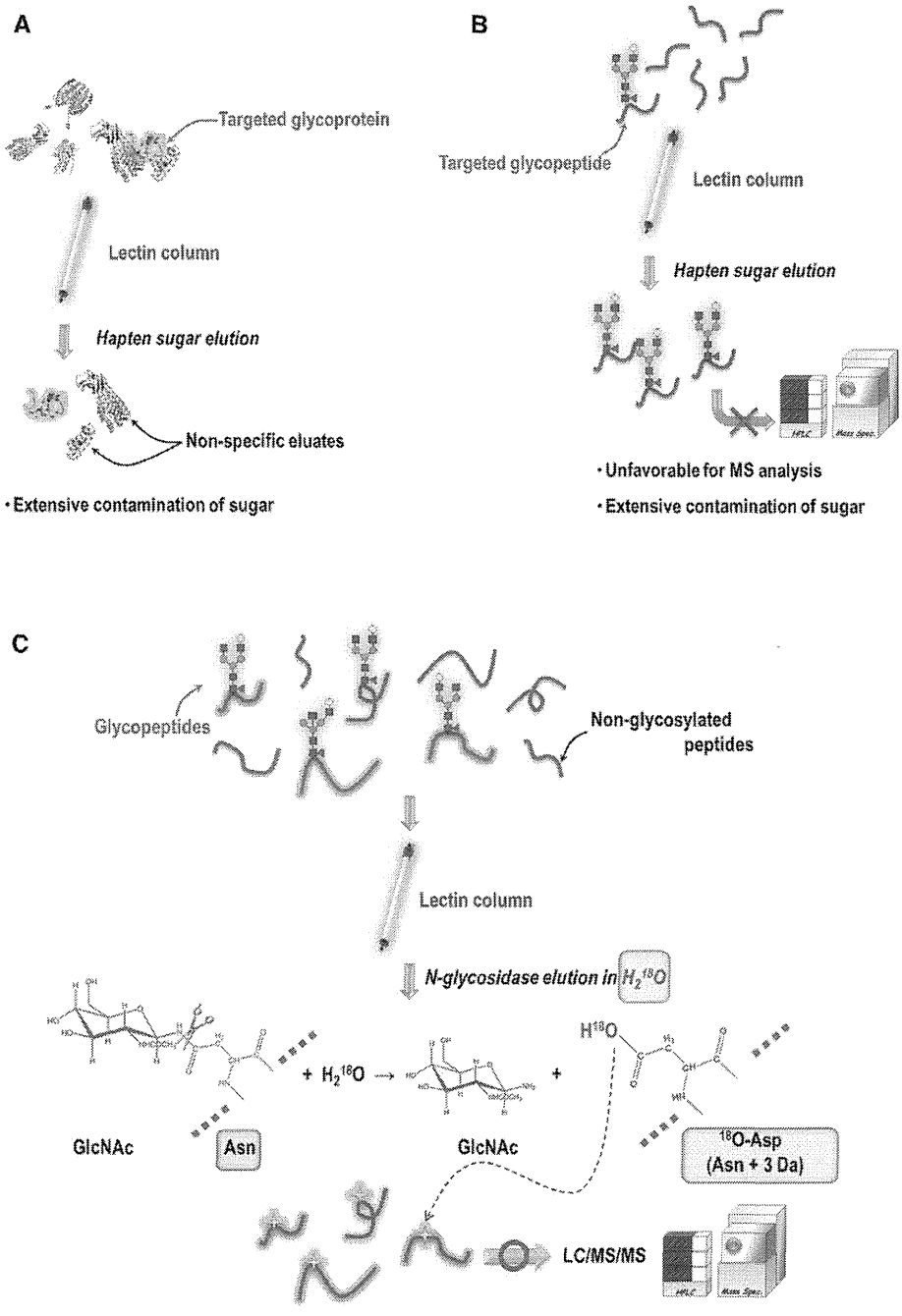


Figure 1. Difficulties and resolution for use of lectin column chromatography in glycoproteomics. (A) Since detergent, high salt, and organic solvent are unusable, lectin column purification using undigested protein samples results in significant contamination of nonspecific proteins due to protein–protein interactions. High concentration of eluting salt is also involved in final product. (B) When starting with digested peptide mixtures, glycopeptide enrichment yield is relatively high. However, direct analysis of eluted glycopeptides is difficult because of low ionization efficiency and impossibility of automated database search. (C) When captured glycopeptides are eluted with *N*-glycosidase, highly specific elution of originally glycosylated peptide could be achieved. If this reaction is combined with ^{18}O-labeling method, direct LC/MS/MS analysis can identify glycosylation sites specifically as 3 Da increased asparagine residues.

2.3 Pre-enrichment of glycopeptides prior to lectin purification

Lectin columns are convenient and widely used enrichment tools for glycoproteins or glycopeptides. As contrasted with the benefits, enrichment ratio itself is not necessarily high enough. When we purified tryptic digest of crude serum by various types of nine different lectin columns according to IGEL method described above, the enrichment efficiency (number of glycopeptide identification / total peptide identification × 100) was only 20–45%. Such insufficient enrichment efficiency was mainly caused by weak lectin–oligosaccharide affinities and the fact that abundance of nonglycosylated peptides in tryptic digest of crude serum samples was absolutely higher than that of glycosylated peptides. This aspect emphasizes the importance of adequate pre-enrichment techniques for pan-glycosylated peptides, such as cellulose column [58–60], graphite carbon column [61–65], and other

hydrophilic affinity resins. Sergei et al. constructed C_{18}-cellulose mix-mode column chromatography and achieved high-yield extraction of *N*- and *O*-glycosylated peptides from mixture of ten standard proteins [59]. Lam et al. applied an online combination of RP/RP and porous graphite carbon LC to the comprehensive analysis of ConA lectin-purified human serum samples and identified 134 *N*-glycosylated serum proteins, 151 possible *N*-glycosylation sites, and more than 40 possible *N*-glycan structures [65]. The CL-4B Sepharose-based hydrophilic extraction of glycopeptides is also one of the most popular techniques in glycoproteomics. Selman et al. loaded 5 μL of CL-4B Sepharose beads into 96-well format plate and enriched human IgG-derived glycopeptides for MALDI-FT-ICR-MS [66]. This type of multiplexed purification system is especially suitable for large-scale biomarker screening assays. Actually, we also applied CL-4B Sepharose beads pre-enrichment procedure before lectin column purification and obtained finally around 90% glycopeptide enrichment ratio [50]. Thus pre-enrichment of total glycopeptides from complex peptide mixtures can drastically improve the glycopeptide focusing efficiency by subsequent lectin column chromatography.

2.4 Selection of lectins for glycoproteomics

So far hundreds of lectins have been isolated from plants, microbes, or animals and most of them are commercially available. The selection of lectins is a critical step for precise and comprehensive profiling of cancerous glycan disorders. In the glycoproteomic studies, both high specificity for glycan structures and high affinity to capture glycopeptides are needed. The glycopeptide enrichment ratio acquired from IGEL purification experiments using human serum and nine distinct lectins (LCA, SNA-I, SNA-II, UEA-I, wheat germ agglutinin (WGA), LPA, ConA, and SSA) was shown in Table 1. Concerning specificity of lectins, LCA, SNA-I, ConA, and SSA demonstrated over 80% glycopeptide enrichment rate, suggesting that ligand specificity of these four lectins would be sufficient. Meanwhile, when looking into numbers of glycopeptide identification, which indirectly reflected the lectin–glycan affinity of each lectin, LCA and SNA-I lectins showed much less glycopeptide recovery rate compared to ConA or SSA. Additionally, the remaining five lectins (Lotus, SNA-II, UEA-I, WGA, and LPA) were scarcely able to capture peptides. At least in our binding condition (100 mM ammonium bicarbonate, 5% ACN, 1 mM calcium chloride, 1 mM manganese chloride), ConA and SSA lectin columns could be considered as appropriate materials to be used for specific and comprehensive profiling of human serum glycoproteome. Individual optimization would be required when analyzing other biological samples or using different condition of binding buffers. In order to cover a larger number of glycan structure changes, it is fundamental to increase options of lectins along with optimum purification protocols.

Table 1. Serum glycopeptide enrichment rate by nine lectin columns

Lectin	Protein IDs[a)]	Peptide IDs[b)]	IGEL(+)[c)]	IGEL (%)[d)]
LCA	109	173	153	88.3
Lotus	51	27	10	36.7
SNA-I	110	194	186	95.9
SNA-II	62	57	33	57.6
UEA-I	13	13	6	46.4
WGA	82	78	41	52.4
LPA	26	14	5	35.7
ConA	183	413	388	93.9
SSA	229	519	425	81.9

a) Number of identified proteins.
b) Number of identified peptides.
c) Number of peptides possessing IGEL tags (3 Da increased asparagine residues).
d) Ratio of IGEL-tagged peptides in total peptide identification, indicating glycopeptide enrichment rate.

In recent studies, multilectin affinity chromatography approaches were developed to enhance glycopeptide recovery rate and expand the comprehensiveness of targeted glycan structures. Zeng et al. combined high abundance protein depletion, ConA-Jacalin-WGA multilectin affinity chromatography, IEF separation, and LC-MS analysis and identified breast cancer associated proteins such as thrombospondin-1 and 5, alpha-1B-glycoprotein, serum amyloid P-component, and tenascin-X, which had potentially abnormal glycans [67]. The same group rigorously evaluated the identical lectin mixture in several glycoproteome profiling studies [68–72]. Qiu et al. integrated serial lectin purification by ConA and SNA with d0- or d3-*N*-acetoxysuccinamide stable isotope labeling on α-amino groups. By use of this methodology, they enabled effective enrichment of sialylated glycopeptides and also differential analysis of those [73].

2.5 Quantitative assessment of glycan structure changes

To identify cancer-associated alterations of glycosylation on multiple proteins, establishment of rigorous quantification strategies should be essential, which could stoichiometrically evaluate the changing rate of each glycoform. Comparative quantification results of only enriched glycopeptides are affected by not only glycan structure changes but also concentration of original core proteins itself, indicating that it is hard to determine whether the identified candidates might be glycosylation-targeting biomarkers or protein concentration biomarkers. Therefore, subtraction of protein concentration effects from quantification results of lectin-enriched glycopeptides is necessary (Fig. 2). Recently, we demonstrated a practical example of this concept for the identification of carbohydrate-targeting lung cancer biomarker discovery [50]. Here we acquired relative quantification profiles from both lectin-purified glycopeptides