Concanavalin A-captured Glycoproteins in Healthy Human Urine*[S]

Linjie Wang‡, Fuxin Li§, Wei Sun‡, Shuzhen Wu‡, Xiaorong Wang‡, Ling Zhang‡, Dexian Zheng‡, Jue Wang§, and Youhe Gao‡¶

Both the urinary proteome and its glycoproteome can reflect human health status, and more directly, functions of kidney and urinary tracts. Because the high abundance protein albumin is not *N*-glycosylated, the urine *N*-glycoprotein enrichment procedure could deplete it, and urine proteome could thus provide a more detailed protein profile in addition to glycosylation information especially when albuminuria occurs in some kidney diseases. In terms of describing the details of urinary proteins, the urine glycoproteome is even a better choice than the proteome itself. Pooled urine samples from healthy volunteers were collected and acetone-precipitated for proteins. *N*-Linked glycoproteins enriched with concanavalin A affinity purification were separated and analyzed by SDS-PAGE-reverse phase LC/MS/MS or two-dimensional LC/MS/MS. A total of 225 urinary proteins were identified based on two-hit criteria with reliability over 97% for each peptide. Among these proteins, 94 were identified in previous urine proteome works, 150 were annotated as glycoproteins in Swiss-Prot, and 43 were predicted as glycoproteins by NetNGlyc 1.0. A number of known biomarkers and disease-related glycoproteins were identified. Because changes in protein quantity or the glycosylation status can lead to changes in the concanavalin A-captured glycoprotein profile, specific urine glycoproteome patterns might be observed for specific pathological conditions as multiplex urinary biomarkers. Knowledge of the urine glycoproteome is important in understanding kidney and body function. ***Molecular & Cellular Proteomics 5:560–562, 2006.***

The urinary proteome has received more and more attention in the proteomic field for its simplicity compared with serum as well as its potential in biomarker discovery. Several research teams have worked on profiling the healthy human urine proteome using electrophoresis and/or liquid chromatography followed by mass spectrometry identification (1–5). The fact that some proteins were observed at higher molecular weight than their theoretical ones on SDS-PAGE (5) confirmed that the post-translational modifications including glycosylation exist extensively in urinary proteins. Glycosylation, a common and important protein post-translational modification, is involved in many biological processes such as cell adhesion, signal transduction, immune response, and inflammatory reaction (6). More than half of all proteins are thought to be glycoproteins, and they will undergo changes in quality and quantity along with the changes in different physiological and pathological states. Both the urine proteome and its glycoproteome can reflect human health status, especially functions of kidney and urinary tracts. In some kidney diseases, albuminuria usually occurs. Because high abundance albumin is not *N*-glycosylated, urine *N*-glycoprotein enrichment procedure could deplete it, and the urine glycoproteome could provide a more detailed protein profile in addition to the glycosylation information. In terms of describing details of urinary proteins, the urine glycoproteome is even a better choice than the proteome itself. In this study our goal was to profile the *N*-linked glycoproteome in normal human urine.

In the present study, pooled urine samples from healthy males and females were collected, and urine proteins were acetone-precipitated as described previously (5). Lectin concanavalin A (Con A)[1] was chosen to enrich *N*-linked glycoproteins for its broader specificity and higher affinity. Con A affinity chromatography was performed according to previous protocols (7). The eluted proteins (100 μg loaded each time) from Con A-agarose were separated and analyzed by two approaches: 1) SDS-PAGE-RPLC/MS/MS (SDS-PAGE, in-gel digestion, and peptide extraction followed by RPLC/MS/MS) and 2) two-dimensional LC/MS/MS (protein mixture digestion followed by strong cation exchange-RPLC/MS/MS). Proteins were reduced with dithiothreitol and alkylated with iodoacetamide before tryptic digestion. Glycosidase PNGase F was also added to remove *N*-glycans from glycoproteins during the digestion. All peptides were analyzed by an LCQ-DECA XP^{plus} electrospray ion trap mass spectrometer (ThermoFinnigan, San Jose, CA). Ions were detected in a survey scan from 400 to 1500 amu (three microscans) followed by five data-dependent MS/MS scans (five microscans each; isolation width, 3 amu; 35% normalized collision energy; dynamic exclusion

From the ‡Proteomics Research Center, National Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing 100005 and §Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China

Received, November 16, 2005

Published, MCP Papers in Press, November 29, 2005, DOI 10.1074/mcp.D500013-MCP200

[1] The abbreviations used are: Con A, concanavalin A; AMASS, advanced mass spectrum scanner; RP, reverse phase.

for 3 min) in a completely automated fashion. Both approaches were run twice in parallel. All MS/MS spectra were searched using Bioworks 3.1 against the database ipi.human.v3.05 (8) with enzyme constraints and with a static modification of +57 Da on cysteine residue and a differential modification of +1 Da on asparagine residue. The precursor ion mass tolerance was 1.40 Da, and the fragment ion mass tolerance was 1.50 Da. We used SEQUEST criteria as follows: $\Delta Cn \geq 0.1$; Rsp = 1; Xcorr $\geq$ 1.9 for +1 charged peptides; Xcorr $\geq$ 2.2 for +2 with fully or partially tryptic end; Xcorr $\geq$ 3.0 for +2 without regard to the end residues; Xcorr $\geq$ 3.75 for +3. Then AMASS version 1.13 (available at www.proteomics-cams.com) was used to filter the SEQUEST results with three parameters: MatchPct $\geq$ 60, Cont $\geq$ 40, and Rscore $<$ 2.6 (9, 10). Proteins with two or more spectra approved by AMASS were accepted as positive identifications. Reverse database searching was used to estimate the false positive rate. The false positive rate = peptide number in reverse database/peptide number in forward database $\times$ 100%, and the final average false positive rate was 2.76% for SEQUEST/AMASS-filtered positive peptides.

In total, 225 proteins were identified (excluding keratins) based on two or more positive peptides with a reliability of more than 97% for each (Supplemental Table 1). For 142 proteins recognized with at least two independent peptides, the reliability of protein identification can reach more than 99.9%. Even for the other 83 proteins identified by single peptide with multiple hits, the reliability was still more than 97%. 94 proteins were identified in previous urine proteome studies (1–5). 43 proteins were also identified in serum *N*-glycoproteome (11–13). 150 were annotated as glycoproteins or subunits of glycoproteins in Swiss-Prot, and 43 were annotated as potential glycoproteins predicted by NetNGlyc 1.0.[2] 22 proteins had potential *N*-linked glycan binding sites but no signal peptides, which means they are unlikely to be *N*-glycosylated. 10 proteins had no *N*-glycosylation sites. Those 32 identified non-*N*-glycosylated proteins might either be associated with the captured *N*-glycoproteins or nonspecifically bind to Con A. For example, albumin can be associated with many proteins in the list as one of its significant functions is protein transportation. The proteins were categorized based on their subcellular localizations. For the known proteins, their subcellular localizations were determined by the Swiss-Prot or Gene Ontology annotations. For all the others, localizations were predicted based on those of similar known proteins by BLAST or PSORT II for proteins that had no similar known proteins in BLAST search. There were 101 extracellular, 67 membrane, 32 lysosomal, 22 cytoplasmic, one cytoskeletal, and two nuclear proteins. They are mainly composed of enzymes, enzyme inhibitors, receptors, immunoglobulin/complements, and apolipoproteins that participate in many biological processes such as immune response and inflammation, blood coagulation, cell adhesion, signal transduction, and cleansing of the aged or abnormal proteins in lysosome. A number of known biomarkers and disease-related glycoproteins were identified, such as prostate-specific antigen, cadherins, and cathepsins. To enrich the details of the urinary Con A-captured glycoprotein profile, 119 proteins identified by one hit were listed in Supplemental Table 2. In total there were 334 proteins identified with more than 97% reliability. The peptide hit numbers of all the proteins were also included in Supplemental Tables 1 and 2 to serve as a rough estimate of the protein quantity in the sample.

Both the changes in protein quantity and the glycosylation status of glycoproteins can be reflected in the profiled changes of Con A-captured glycoproteins. Specific patterns might be observed in the Con A-captured glycoproteins for specific pathological conditions as "multiplex urinary biomarkers." Hence we believe profiling Con A-captured glycoproteins in healthy human urine will be a very useful reference for future applications of the urine glycoproteome. We expect that with the development of more efficient enrichment techniques and sophisticated mass spectrometers more glycoproteins will be in the detectable range.

Acknowledgments—We thank Dr. Mark A. Knepper (Laboratory of Kidney and Electrolyte Metabolism, NHLBI, National Institutes of Health, Bethesda, MD) for helpful discussion. We also thank Dr. Weizhi Chen and Peter Marshall for critical reading of the manuscript.

* This work was supported in part by National Basic Research Program Grant 2004CB520804 and by National Natural Science Foundation Grants 30270657, 30230150, and 3037030. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

¶ To whom correspondence should be addressed: Inst. of Basic Medical Sciences, Chinese Academy of Medical Sciences/Peking Union Medical College, 5 Dong Dan San Tiao, Beijing 100005, China. Tel.: 86-010-6521-2284; Fax: 86-010-6521-2284; E-mail: gaoyouhe@pumc.edu.cn.

REFERENCES

1. Spahr, C. S., Davis, M. T., McGinley, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Courchesne, P. L., Chen. K., Wahl, R. C., Yu, W., Luethy, R., and Patterson, S. D. (2001) Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry: I. Profiling an unfractionated tryptic digest. *Proteomics* **1,** 93–107
2. Davis, M. T., Spahr, C. S., McGinley, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Yu, W., Luethy, R., and Patterson, S. D. (2001) Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry: II. Limitations of complex mixture analyses. *Proteomics* **1,** 108–1171
3. Thongboonkerd, V., McLeish, K. R., Arthur, J. M., and Klein, J. B. (2002) Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation. *Kidney Int.* **62,** 1461–1469
4. Pieper, R., Gatlin, C. L., McGrath, A. M., Makusky, A. J., Mondal, M., Seonarain M., Field, E., Schatz, C. R., Estock, M. A., Ahmed, N., Anderson, N. G., and Steiner, S. (2004) Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct

[2] R. Gupta, E. Jung, and S. Brunak, manuscript in preparation.

protein spots. *Proteomics* **4,** 1159–1174
5. Sun, W., Li, F., Wu, S., Wang, X., Zheng, D., Wang, J., and Gao, Y. (2005) Human urine proteome analysis by three separation approaches. *Proteomics* **5,** 4994–5001
6. Haltiwanger, R. S., and Lowe, J. B. (2004) Role of glycosylation in development. *Annu. Rev. Biochem.* **73,** 491–537
7. Kristiansen, T. Z., Bunkenborg, J., Gronborg, M., Molina, H., Thuluvath, P. J., Argani, P., Goggins, M. G., Maitra, A., and Pandey, A. (2004) A proteomic analysis of human bile. *Mol. Cell. Proteomics* **3,** 715–728
8. Kersey, P. J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E., and Apweiler, R. (2004) The International Protein Index: an integrated database for proteomics experiments. *Proteomics* **4,** 1985–1988
9. Sun, W., Li, F., Wang, J., Zheng, D., and Gao, Y. (2004) AMASS: software for automatically validating the quality of MS/MS spectrum from SEQUEST results. *Mol. Cell. Proteomics* **3,** 1194–1199
10. Li, F., Sun, W., Gao, Y., and Wang, J. (2004) Rscore: a peptide randomicity score for evaluating tandem mass spectra. *Rapid Commun. Mass Spectrom.* **18,** 1–5
11. Yang, Z., and Hancock, W. S. (2004) Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J. Chromatogr. A* **1053,** 79–88
12. Bunkenborg, J., Pilch, B. J., Podtelejnikov, A. V., and Wisniewski, J. R. (2004) Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* **4,** 454–465
13. Yang, Z., Hancock, W. S., Chew, T. R., and Bonilla, L. (2005) A study of glycoproteins in human serum and plasma reference standards (HUPO) using multilectin affinity chromatography coupled with RPLC-MS/MS. *Proteomics* **5,** 3353–3356