research articles Journal of proteome research

Lectin-Based Enrichment Method for Glycoproteomics Using Hollow Fiber Flow Field-Flow Fractionation: Application to *Streptococcus pyogenes*

Dukjin Kang,[†] Eun Sun Ji,[†] Myeong Hee Moon,[*,‡] and Jong Shin Yoo[*,†]

Mass Spectrometry Research Center, Korea Basic Science Institute, Ochang, Cheongwon-Gun, Chungcheongbuk-Do, 363-883, Korea, and Department of Chemistry, Yonsei University, Seoul, 120-749, Korea

Received October 16, 2009

This paper presents a new application of hollow fiber flow field-flow fractionation (HF5) as a preparative method to preconcentrate high mannose type N-linked glycoproteins from *Streptococcus pyogenes* by means of the mannose-specific binding affinity between concanavalian A (ConA) and N-linked glycosylated proteins. Prior to fractionation of N-linked glycoproteins from bacterial lysates, it was examined that ConA formed several types of multimers depending on the pH values (4, 6, and 8) of the carrier solution and it was confirmed that the molecular weight (MW) of ConA, spiked with α-1 acid glycoprotein (AGP) as a standard glycoprotein, increased due to binding with the mannose moiety of AGP. After adding ConA to bacterial lysates, mannose type N-linked glycoproteins were found to be enriched when the ConA fraction was isolated from whole bacterial lysates through HF5 run. For the identification of glycoproteins, the ConA fraction of HF5 was tryptically digested and followed by two-dimensional nanoflow strong cation exchange-reversed phase liquid chromatography-electrospray ionization-tandem mass spectrometry (2D SCX-RPLC-ESI-MS-MS) analysis to identify the N-linked glycoprotein species. From two-dimensional shotgun analyses, 45 proteins that exist on the Asn-Xaa-Ser/Thr sequence were identified as high mannose type N-linked glycoprotein. As a result, it was first demonstrated that HF5 is an alternative tool to enrich high mannose type N-linked glycoproteins using ConA-specific binding affinity.

Keywords: Hollow Fiber Flow Field-Flow Fractionation • Glycoproteomics • N-linked Glycoprotein • 2D SCX-RPLC-ESI-MS-MS

Introduction

Advanced mass spectrometry (MS)-based proteomics in the determination of post-translational modifications (PTMs) and their structural compositions upon specific amino acid sites of proteins has become an important tool for elucidating complicated proteomic mechanisms, related to the regulation of protein activity and interactions, and the delivery of target signals to specific organelles/membranes in a cell.[1−3] Pathogenically, glycosylation, one of PTMs, has also been required for unveiling clinical determinants for diverse diseases (e.g., immune deficiencies, neurodegenerative diseases, cardiovascular diseases, and cancer).[3−5] Furthermore, the quantitative/qualitative analyses of the diverse compositions of carbohydrate species that are covalently attached to the glycoproteome provide possible therapeutic solutions for early diagnosis by excavating glycoprotein biomarkers.[3,6,7] Although the identification of a glycoprotein and its carbohydrate structure is clinically important to elucidate the virulence factors of various diseases, it is not simple due to the low abundance of glycopeptides obtained from tryptic cleavage, which comprise approximately about 2−5% of the total amount of peptides. Moreover, it is not straightforward to characterize glycosylated peptides in MS experiments since ionization of glycosylated peptides containing a number of glycans is relatively poor and the carbohydrate linkages are preferentially cleaved during CID with incomplete fragmentation of peptide backbone.[8,9]

To overcome these limitations, various types of lectin-based affinity chromatographic methods have been used to selectively isolate glycoproteins from several thousands of protein mixtures for enhancing the efficiency of identifying glycoproteins and their glycan moieties.[3,6] Lectins are particularly capable of selective binding with carbohydrate species, such as mannose, *N*-acetylglucosamine, *N*-acetylgalactosamine, *N*-acetylneuraminic acid, and fucose. Concanavaline A (ConA), one of the lectins, is universally used for the enrichment of N-linked glycoproteins, because of its high affinity for mannose type N-linked glycoprotein accompanied with metal (calcium, magnesium, and manganese) ions.[13] Recently, several types of ConA immobilized forms, such as cartridge, colloidal gold, membrane, and monolith column, have provided the feasibility for

* To whom correspondence should be addressed. Jong Shin Yoo, Mass Spectrometry Research Center, Korea Basic Science Institute, Ochange, Cheongwon-Gun, Chungcheongbuk-Do, 363-883, Korea. Phone, (82) 43 240 5150; fax, (82) 43 240 5159; e-mail, jongshin@kbsi.re.kr. Myeong Hee Moon, Department of Chemistry, Yonsei University, Seoul, 120-749, Korea. Phone, (82) 2123 5634; fax, (82) 2 364 7050; e-mail, mhmoon@yonsei.ac.kr.

[†] Korea Basic Science Institute.

[‡] Yonsei University.

10.1021/pr900937w © 2010 American Chemical Society

Published on Web 04/08/2010

conventional analytical systems to preconcentrate glycoproteins from biological fluid and organs.[10,11,13] These conventional tools, however, have several drawbacks such as the limited concentration of the protein mixtures, and low sample recovery and reproducibility.

Hollow fiber flow field-flow fractionation (HF5), one of the flow field-flow fractionation (FlFFF or F4) variants, has been widely used for separating biological components such as proteins, DNA, and cells.[16−20] HF5 was utilized with MS for protein characterization either by off-line matrix assisted laser desorption ionization/time-of-flight mass spectrometry (MALDI/TOFMS) applied for whole bacteria,[21] recombinant proteins drugs,[22] and blood serum,[23] or by online electrospray ionization/TOFMS (ESI/TOFMS) for intact proteins.[24] Very recently, nongel based two-dimensional separation methods were introduced by Moon and colleagues by coupling FlFFF with isoelectric focusing (IEF) technique in order to overcome drawbacks of conventional gel-based techniques (e.g., 1D- or 2D-PAGE) that are labor-intensive, time-consuming, and cause denaturation of protein degradation.[25,26] The combined IEF−FFF techniques were applied to fractionate intact and low-abundance proteins from the human urinary proteome according to their isoelectric point (p*I*) and hydrodynamic size (d_s), and the fractionated proteins were characterized by nanoflow liquid chromatography/ESI/tandem mass spectrometry (nLC-ESI-MS-MS). Furthermore, FlFFF has also been examined as an alternative tool to complement conventional density gradient centrifugation methods for characterizing size-dependent proteome patterns of biological vesicles such as mitochondria,[27] exosomes,[28] and membrane proteins.[29] Differing from traditional FlFFF channels, in which separation of the protein is carried out by manipulating the application of cross-flow driven at the direction perpendicular to the migration flow in a flat rectangular membrane, separation in HF5 can be achieved by controlling the radial flow rate as the driving force of sample retention toward the inner wall of a hollow fiber (HF). When proteins are loaded into one end of the HF, they are differentially distributed from the inner wall of the HF in compliance with the radial equilibrium positions that are dependent on the counterbalancing of two forces between radial flow and diffusion. Because of the differences in diffusion coefficients of proteins, smaller sized proteins will be elevated further away from the inner wall of the HF than larger ones. Consequently, when a laminar flow is applied along the fiber axis for sample migration, the smaller proteins that are entrained in the faster laminar flow streamline are eluted earlier than the larger ones, and thus, separation is achieved in an increasing order of diffusion coefficients or hydrodynamic diameters. In the enlarged diagram of Figure 1, the separation mechanism of HF5 is illustrated and it is also detailed in our previous studies.[18,30] Especially, the use of micrometer-sized HF (450 μm i.d.) has provided several merits: (i) the separation of proteins and nanosized particles in the microflow rate scale which can reduce injection amounts, (ii) a minimization of dead volume using a glue-free connection between both ends of the HF and the inlet/outlet tubings, and (iii) the feasibility of integration with different types of separation or detection technique such as electrospray ionization.

In this study, HF5 was used as a promising new tool for preconcentration of high mannose type N-linked glycoproteins from bacterial lysates as a test model by means of ConA-specific affinity without relying on the conventional column based lectin affinity techniques. Prior to the enrichment of N-linked glycoproteins, pH of the carrier solutions were adjusted to make multimeric forms of ConA so that N-linked glycoproteins bound to ConA can be separated from free cytoplasmic proteins due to the differences in retention times in HF5 run. Moreover, the affinity of ConA was also examined using a α1-acid glycoprotein (or orosomucoid) as a standard glycoprotein for optimizing the molar ratio of ConA versus AGP. Biologically, AGP is composed of approximately 45% carbohydrate attached in the form of five complex-type N-linked glycans and can express any of the glycans corresponding to different forms of branching such as di-, tri-, or tetra-antennary glycans.[14,15] The potential of HF5 toward glycoproteins was evaluated by fractionating mixtures of ConA and bacterial lysates of *Streptococcus pyogenes* and the eluted ConA bound with glycoproteins were collected for the secondary examination. The N-linked glycoproteins that were immobilized to ConA were tryptically digested, and followed by online two-dimensional strong cation exchange-reversed phase-reversed phase liquid chromatography-electrospray ionization-tandem mass spectrometry (2D SCX-RPLC-ESI-MS-MS) for glycoprotein identification. The experimental workflows for the enrichment of mannose type N-linked glycoproteins from *S. pyogenes* using HF5 are illustrated in Scheme 1.

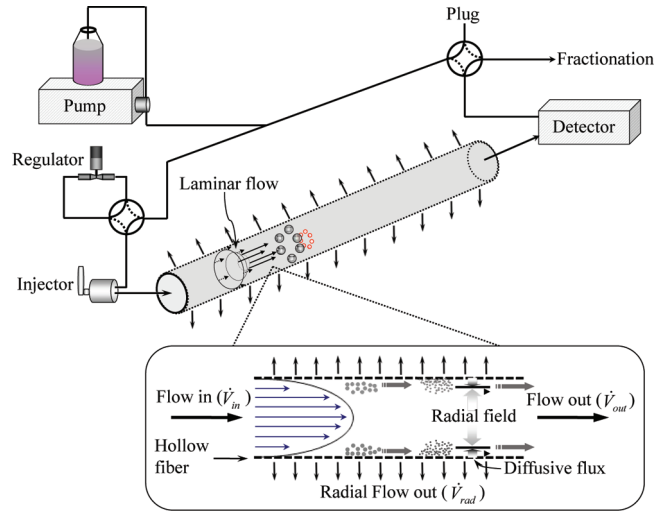


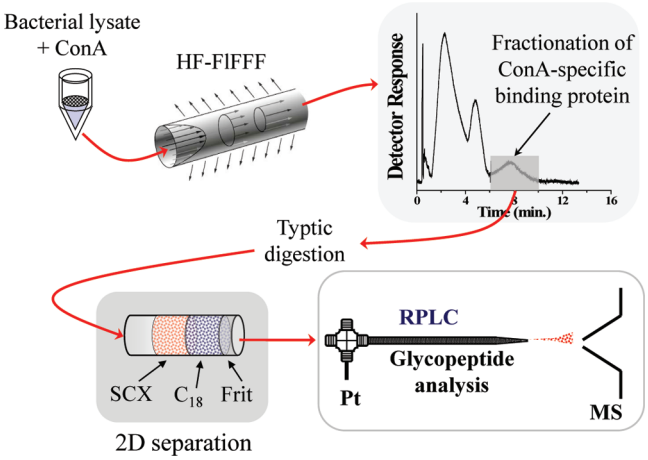
Figure 1. The system configuration of hollow fiber flow field-flow fractionation (HF5). After focusing/relaxation step (dotted line), separation of protein standards and bacterial lysates by HF5 were achieved according to their sizes or MWs through switching back to a solid line connection. The enlarged side view of the HF module illustrates separation mechanism in the HF module.

Materials and Methods

Culture of *S. pyogenes*. A *S. pyogenes* M1 strain was cultured in Todd−Hewitt broth (BD) supplemented with 0.5% yeast extract (THY) in 5% CO_2 at 37 °C. Bacterial growth was monitored by measuring the absorbance at 600 nm using a Beckman DU 800 spectrophotometer. Colonies were grown on Todd−Hewitt agar plates and then were resuspended in 500 mL of THY broth and incubated to an optical density of 0.9 at 600 nm. Harvested cells were suspended in 20 mM Tris-HCl buffer (pH 8.0), and disrupted twice by a French pressure (SLM AMINCO) at 20 000 lb/in.^{2}, The bacterial debris was removed by centrifugation at 15 000*g* for 20 min at 4 °C. The bacterial lysates retrieved from the centrifugal method were stored at −80 °C before enzymatic cleavage.

Hollow Fiber Flow Field-Flow Fractionation (HF5). The HF5 system used for the enrichment of N-linked glycoproteins from

Scheme 1. The Entire Experimental Procedure for Mannose Type N-Linked Glycoproteomic Analysis of Bacterial Lysates of *S. pyogenes* by Using an Affinity for ConA and Molecular Mass Sorting by HF5



the bacterial lysates by means of retrieving the ConA fraction through HF5 run was constructed in our laboratory. The entire configuration of the HF5 setup was detailed in our previous studies.[18,26,30] Briefly, the HF used in this study was made of polysulfone (MW cutoff of 30 kDa) that was obtained from Kolon Central Research Institute (Yongin, Korea). The HF used was 25-cm in length, 450-μm i.d., and 750-μm o.d. The HF module was constructed by inserting a single hollow fiber into glass tubing (3.2-mm o.d. and 1.6-mm i.d.), and then both ends of the HF were connected with a fused silica capillary (360-μm o.d., 100-μm i.d.) by inserting a few millimeters of the capillary tubing into each end of the HF, which was compassed with a piece of Teflon sleeve (1/16 in.) from Upchurch Scientific, Inc. (Oak Harbor, WA). The connection of each end of the HF with the glass housing was fastened by using a union between a 1/8-in. ferrule with a 1/8-in. male nut for the glass tubing and a 1/16-in. ferrule with a 1/16-in. male nut for the HF. The injector for introduction of proteins into the HF and the regulator for manipulation of the focusing/relaxation step were connected between the HF inlet and HPLC pump. The HF5 setup is illustrated in Figure 1. The fittings used in this study were purchased from Upchurch Scientific, Inc. (Oak Harbor, WA).

To optimize the performance of HF5 in the enrichment of ConA-specific binding proteins from bacterial lysates, the following protein standards were used: concanavalian A, carbonic anhydrase (29 kDa), alcohol dehydrogenase (150 kDa) from yeast (YADH), and apoferritin (444 kDa) from Sigma (St. Louis, MO). The carrier solution used for HF5 run was composed of 50 mM Tris-HCl solution (pH 8.0) containing 150 mM sodium chloride, 1 mM magnesium chloride, 1 mM manganese chloride, 1 mM calcium chloride, and 0.02% sodium azide from Sigma (St. Louis, MO). Prior to use, all buffer solutions used in this study were prepared from ultrapure water (>18 MΩ) and filtered with a membrane having a pore size of 0.45 μm from Millipore Corp. (Bedford, MA). The delivery of the carrier solution to the HF module and the injection of sample components were carried out using an Agilent 1090 system (Palo Alto, CA). The protein separation by HF5 was carried out by the following steps: (i) focusing/relaxation, and (ii) elution. At the focusing/relaxation step, the pump flow was split into two steams with the ratio of 1:9 by adjusting a metering valve and two four-way valves (dotted line of the four-way valve in Figure 1). One (1/10 of the total flow rate) of the two streams entered into the HF inlet through an autosampler, and the rest (9/10 of the total flow rate) entered through the HF outlet via a UV detector. During this process, the sample components, such as protein standards and bacterial lysates, were introduced into the HF inlet by entraining the flow stream (1/10 of the total flow rate), and followed by focusing/relaxation in which sample components were focused at a position 1/10 of fiber length with equilibrium between diffusive force and the force of radial flow. After focusing/relaxation step, the streamlines of both valves were switched back to a solid line connection (solid line of the four-way valve in Figure 1) to carry out the elution step. At the elution step, the separation of proteins in the HF was accomplished by controlling both the radial flow rate and the outflow rate with a capillary tubing (360-μm o.d., 75-μm i.d.) located at the outflow outlet. The detection of proteins eluted from the HF was achieved by using a UV detector (Agilent 1090 system) at a wavelength of 280 nm. For the purpose of enriching ConA-specific binding glycoproteins from the bacterial lysates, 5 μL of ConA (1 μg/μL) was added to 95 μL of lysates (1 μg/μL) from *S. pyogenes* and then, 100 μL (10 $\times$ 10 for single) of lysates spiked with ConA was injected into the HF for each run. The ConA fraction of the HF5 was concentrated and sequentially stored at −4 °C before enzymatic treatment.

Tryptic Digestion. Prior to tryptic digestion, the ConA fraction of the HF5 was purified to eliminate the metal ions contained in the carrier solution by using a 3 kDa MWCO Microcon (YM-3) centrifugal filter from Millipore (Danvers, MA) and then, sequentially washed with ice-cold 10 mM ammonium bicarbonate (pH 8) solution. The concentration of the ConA fraction was examined by the Bradford assay and followed by which they were digested simultaneously by trypsin. After reducing the volume of the ConA fraction to be about 10 μL with a SpeedVac, it was spiked into the denaturing buffer solution (8 M urea, 10 mM dithiothreitol, and 0.1 M PBS) for 2 h at 37 °C. To alkylate the remaining thiol group, the solution was treated with 20 mM iodoacetamide (IAM) in an ice bath for 2 h in the dark. Cysteine was then added to the solution to remove IAM, and the solution was allowed to remain at room temperature for 30 min. The solution was diluted to a final concentration of 1 M Urea with 0.1 M phosphate buffer. Tryptic digestion was carried out by adding a sequencing graded modified trypsin from Promega Corp. (Madison, WI) to the diluted solution in the ratio of 1:40 (trypsin/protein) and subsequently incubating at 37 °C for 18 h. To stop the tryptic digestion, the solution was treated with *N*-α-*para*-tosyl-L-lysine chloromethylketone hydrochloride (TLCK) in slight molar excess to trypsin to deactivate trypsin. After proteolytic cleavage, the digested protein solution was desalted using an Oasis HLB cartridge from Waters (Milford, MA) and the volume of the solution was reduced to approximately 20 μL using SpeedVac.

Two-Dimensional NanoLC-ESI-MS-MS. Two-dimensional nanoflow strong cation exchange-reversed phase liquid chromatography-electrospray ionization-tandem mass spectrometry (2D SCX-RPLC-ESI-MS-MS) was carried out by a microflow U3000 HPLC system from Dionex (Sunnyvale, CA) coupled to 7 T LTQ Fourier transform ion cyclotron resonance mass spectrometer (7T LTQ-FTICR-MS) from Thermo Finnigan (San Jose, CA) via electrospray ionization (ESI). Peptide mixtures digested from the ConA fraction were separated using online

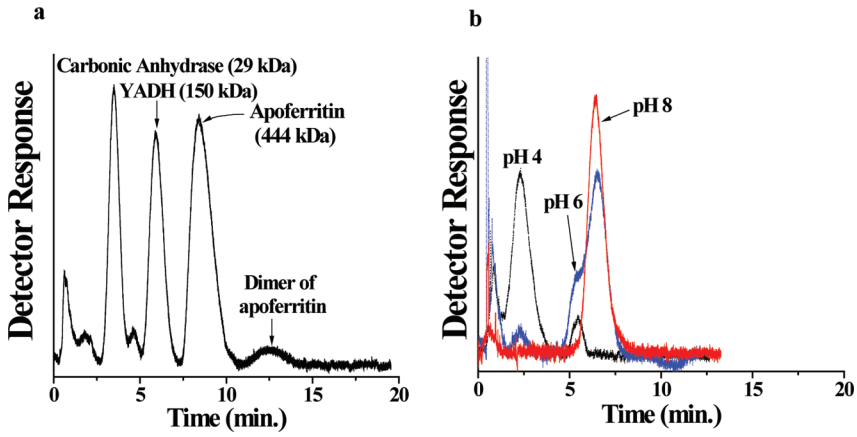


Figure 2. (a) The fractogram of protein standards, such as carbonic anhydrase (29 kDa), yeast alcohol dehydrogenase (150 kDa), and Apoferritin (444 kDa), was obtained at the flow rate conditions: outflow and radial flow rates for HF5 run were 40 μL/min and 0.46 mL/min, respectively. (b) The effect of pH of carrier solution on elution profile that represents the various multimeric structures of ConA at the three different pH values of 4.0 (black), 6.0 (blue), and 8.0 (red).

two-dimensional nanoflow strong cation exchange-reversed phase liquid chromatography (SCX-RPLC), which consisted of a dual purpose sample trap (SCX and RP resins) with an analytical capillary RPLC column (170 mm $\times$75 μm). Details of this procedure can be found in our previous studies.[29,31,32] A homemade pulled tip analytical capillary column used in this study was prepared by packing with Magic C_{18}AQ (3 μm, 100 Å) resins from Michrom BioResources, Inc. (Auburn, CA). For the purpose of online 2D SCX-RPLC separation for peptide mixtures, the dual trap column was prepared in a capillary (200 μm-i.d., 360 μm-o.d.) with an end frit (2 mm in length) by the sol−gel polymerization method at one end of capillary, and the capillary was packed in sequence, first with Magic C_{18}AQ (5 μm, 200 Å) resins for 0.5 cm and then with Polysulfethyl A SCX resin (5 μm, 300 Å) from The Nest Group, Inc. (Southboro, MA) for 1.5 cm. The dual trap column and the analytical column were connected via a PEEK microcross as shown in Scheme 1.

For 2D separation, six salt steps of 0, 2, 5, 10, 50, and 1000 mM NH_4HCO_3 solutions were used to fractionate the peptide mixture from the SCX trap to the RP part of the dual trap column. After each salt step, the fractionated peptides bound to the RP part were separated via a binary gradient RPLC run. For the RPLC runs, two buffer solutions, (A) 2% acetonitrile (ACN) and (B) 95% ACN in water, both with 0.1% formic acid, were used as mobile phase solutions. The flow rate for RPLC run was 200 nL/min and the binary gradient run conditions were as follows: buffer B increased from 5 to 10% for 12 min, then linearly increased to 18% over 18 min, to 32% over 55 min, ramped to 80% over 3 min, was held at 80% for 10 min to clean the RP column, and finally, it was decreased to 5% over 2 min, and was held for 20 min to re-equilibrate the RP column. The eluted peptides from the capillary column were directly fed into the 7T LTQ-FTICR-MS via ESI in positive ion mode and a voltage of 2.0 kV was applied through the Pt wire connected to the microcross. MS analysis was carried out by each precursor scan (300−1800 amu) followed by three data-dependent MS/MS scans. The MS/MS spectra were analyzed using a Mascot Search program against a bacterial database. The mass tolerance between the measured monoisotopic mass and the calculated mass was 0.5 u for the molar mass of a precursor peptide and 0.2 u for the mass of peptide fragment ions. On the base of the database searching algorithm, only proteins were considered acceptable in the case of satisfying the condition in which those were matched with more than two peptides, which was larger than a minimum Mascot score of 43. The fixed modification was set as carbamidomethylation of cysteine including variable modification such as oxidation of methionine. tryptic enzyme, and double miscleavage was defined.

Results and Discussion

As a preparative tool, the HF5 was used to preconcentrate N-linked glycoproteins from protein mixtures by means of an intrinsic affinity of ConA for a high mannose type N-linked glycoprotein without the complicated preparative steps that are accompanied with conventional column based techniques. Additionally, MW increase (over ~204 kDa) of ConA through the formation of multimers, determined by the pH value of the carrier solution, was suitable for isolating N-linked glycoproteins from free cytoplasmic proteins. For the purpose of enriching N-linked glycoproteins from lysates of *S. pyogenes*, a 10 μL aliquot of bacterial lysate that was prepared by mixing 95 μL (1 μg/μL) of the lysate with 5 μL (1 μg/μL) of ConA was injected to the HF5, and it was repeated 10 times for the accumulation of fractions to be collected. During HF5 runs, the ConA fraction was digested for proteolytic cleavage and the resulting peptide mixtures were analyzed by 2D SCX-RPLC-ESI-MS-MS to identify ConA-specific binding proteins.

Prior to fractionation of the ConA-specific binding glycoproteins, performance of HF5 separation was evaluated using three protein standards, carbonic anhydrase (29 kDa), YADH (150 kDa), and apoferritin (444 kDa), and the results were utilized to estimate the MW distribution of ConA in relation to the variation of pH values of the carrier solution. In Figure 2a, the fractogram shows that the three protein standards were clearly separated in an increasing order of their diffusion coefficients or their MWs. The fractogram separating the three proteins was obtained at outflow rate of 40 μL/min and a radial flow rate of 0.46 mL/min. Interestingly, the two peaks of YADH (150 kDa) and apoferritin (444 kDa), consisting of 4 subunits and 23−25 subunits (most likely 24), respectively, were eluted corresponding with their intact structures without being dissociated. If they were dissociated for some reasons, their retention times would have been shifted to shorter time scales.

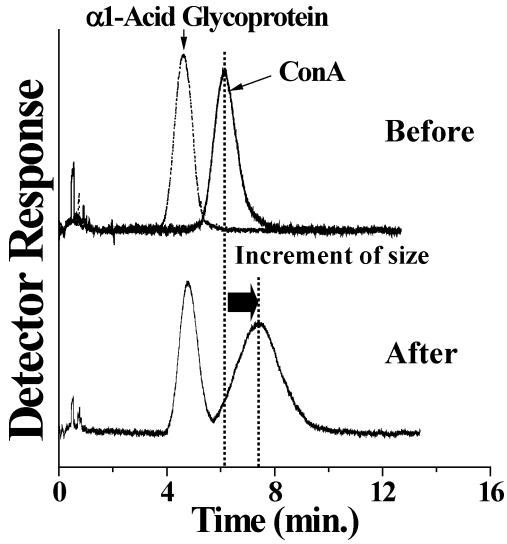


Figure 3. HF5 separation of α1-acid glycoprotein (AGP) and concanavalian A (ConA) (top) before and (bottom) after mixing each other. The top fractograms were achieved by individual injections and the bottom fractogram shows the affinity of AGP to ConA. Flow rate conditions are the same as used in Figure 2a.

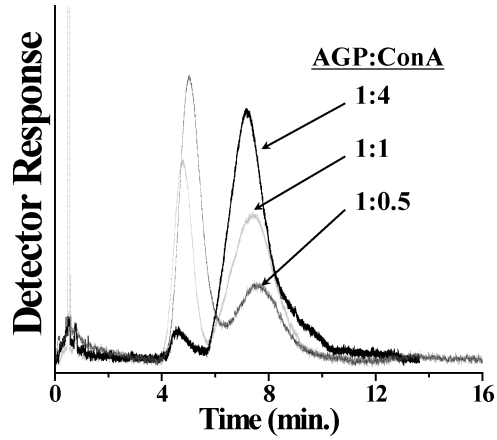


Figure 4. The superimposed fractograms of AGP and AGP bound ConA obtained by varying mass ration of AGP/ConA as 1:0.5, 1:1, and 1:4. Injection amount of AGP was fixed as 1 μg except for same ratio 1:0.5 (2 μg of AGP was used in this case, but the entire fractograms were area-normalized to compare the relative distribution of each component).

This demonstrates that HF5 is a suitable method for separating intact proteins without the dissociation of subunit-assembled proteins. For fractionation of N-linked glycoproteins via enrichment of the ConA fraction in HF5 run, the MW distributions of ConA were evaluated depending on the pH values of the carrier solution. While ConA exists as a tetramer having a MW of approximately 104 kDa under neutral pH, it exhibits a multimer or aggregated forms above pH 8.0 and it dissociates into dimers and monomers below pH 5.0. In Figure 2b, ConA appeared as various multimeric moieties depending on the three different pH values of 4.0, 6.0, and 8.0. At pH 4.0, two peaks were observed at 2.3 and 5.4 min. It is presumed that the first peak can be the dissociated monomers about 26 kDa in accordance with the MW distribution of the protein standards in Figure 2a. At pH 8.0, ConA uniquely formed the octameric structure as consisted of two tetramers of about 208 kDa as a total size, whereas both the octamer and tetramer existed together at pH 6.0. From the examination of pH influence of the carrier solution on multimeric formation of ConA, it was found that the octamerized ConA is more suitable for isolating a ConA-specific binding N-linked glycoprotein from free cytoplasmic proteins, since most of the cytoplasmic proteome generally has a MW distribution of below ~100 kDa, except for some membrane proteins from the cell wall or subcellular organelles.

To examine the affinity of ConA for N-linked glycoproteins by size separation with HF5, α1-acid glycoprotein (AGP) was spiked into the ConA solution and the mixture was injected to the HF module in order to evaluate whether AGP has an affinity to ConA. The top fractogram in Figure 3 was obtained by injecting ConA and AGP individually without mixing, showing that AGP and ConA were eluted at different retention times of 4.8 and 6.2 min, respectively, according to their intact MWs. When it is tested with ConA mixed with AGP shown at the bottom of Figure 3, elution time of ConA bound with AGP increased to 7.3 min from 6.2 min with a broader MW distribution while AGP maintained the same peak shape as observed in individual run. On the basis of the relationship between MW values and the retention time of the protein standard in Figure 2a, it could be expected that the MW distribution of AGP bound ConA dramatically increased in the range of about 200−400 kDa. Standard deviations in retention time and peak area from three repeated runs are 7.9% and 1.6%, respectively. From this result, it is evident that the HF5 can be a promising tool for isolating mannose type N-linked glycoproteins from nonglycosylated ones by size based sorting. Additionally, we also examined the binding efficiency of ConA with AGP by varying mixing ratio of ConA to AGP, since the relative percentage of mannose in total glycans can be a critical factor in the determination of a specific affinity for ConA. In the case of AGP, about ~20% of the mannose exist as of di-, tri-, and tetra-glycan structures. The fractogram in Figure 4 was obtained at three different ratios of AGP/ConA as 1/0.5 (3 μg), 1/1 (2 μg), and 1/4 (5 μg). For the case of the mass ratio of 1:0.5, a total 3.0 μg was injected instead of 1.5 μg since the UV detector used in this study is not suitable for detecting the ConA (0.5 μg which is nearly 200 fmol) due to the limitation of detection. Each fractogram obtained by HF5 in Figure 4 was normalized in order to compare the relative population of unbound AGP and ConA−AGP. As shown in Figure 4, when the ratio of AGP/ConA is adjusted to 1:4, most of AGP appeared as bound to ConA. Interestingly, the molar ratio of ConA needed to completely bind with AGP is close to the occurrence percentage (about ~20%) of mannose out of total glycans in AGP. From this result, it is presumed that the number of moles of ConA required is dependent on the percentage of mannose contained in glycan structures. On the other hand, when the mixing ratios of AGP/ConA are 1:0.5 and 1:1, unbound AGP eluted in intact form.

To verify the potential of HF5 for glycoproteomics, bacterial lysates from *S. pyogenes* were utilized for the study of enrichment of N-linked glycoproteins with ConA. Prior to the fractionation of N-linked glycoproteins, bacterial lysates containing ConA were prepared. A total volume of 100 μL was prepared by adding 5 μg of ConA to 95 μg of the lysate. Then, a 10 μL aliquot of the lysates was injected into the HF channel for the collection of ConA bound with glycoproteins and it was repeated by 10 batch injections to accumulate fractions for further examination of glycoproteins. In Figure 5, the fracto-

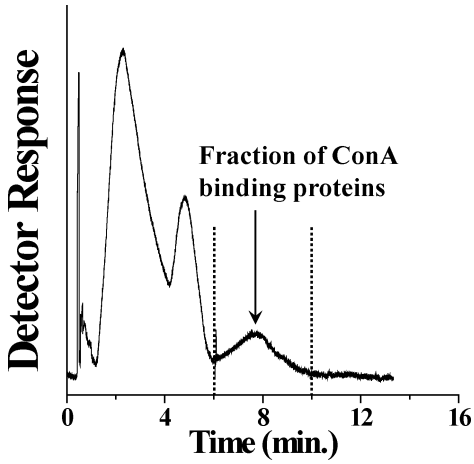


Figure 5. The HF5 fractogram of a 10 μL aliquot of the mixture of lysate and ConA (5 μg of ConA to 95 μg of the lysate in total 100 μL). The fraction of ConA−glycoprotein complex (6−10 min.) was collected for 10 repeated batch injections.

gram of the ConA−lysate mixtures shows three peaks: the elution of free cytoplasmic proteins appeared with bimodal distribution (1−6 min) and the elution of ConA bound with glycoproteins (6−10 min). Especially, based on MW distribution of protein standards in Figure 2a, it is confirmed that a majority of free cytoplasmic proteins (1−6 min) were less than 100 kDa, whereas the MWs of ConA (from 6−10 min) were more broadened throughout 100−400 kDa, due to the binding of various N-linked glycoproteins to ConA. To identify the enriched N-lined glycoproteins, the ConA−glycoproteins fraction (6−10 min) collected from repeated HF5 runs was digested with trypsin, and then resulting peptide mixtures were analyzed by online 2D SCX-RPLC-ESI-MS-MS for protein identification. Peptide mixtures from ConA−glycoproteins fraction were two-dimensionally separated in a trapping column packed with SCX and RP resin prior to analytical column at six salt steps of increasing concentrations of ammonium bicarbonate. After each salt step displacement, a binary gradient separation using RPLC-ESI-MS was carried out. The MS/MS spectra in Figure 6 were obtained through collision induced dissociation (CID) for a precursor ion m/z 719.37 $[M + 2H^+]^{2+}$ detected at the retention time of 51.1 min during RPLC-ESI-MS run of the digested peptide mixtures from the ConA−glycoproteins fraction. The RPLC run shown in Figure 6 was made right after the salt step elution at 2 mM NH_4HCO_3 (each chromatogram of RPLC run was not included here for each salt step). A database search against the bacterial database resulted in the peptide sequence K.FEGTEDGVETILK.A, which was derived from *Enolase*. This protein is well-known as an essential component for the degradation of carbohydrates via glycolysis. It is known to allow the bacterium to acquire surface-associated proteolytic activity, which in turn contributes to tissue invasion and virulence.[33,34]

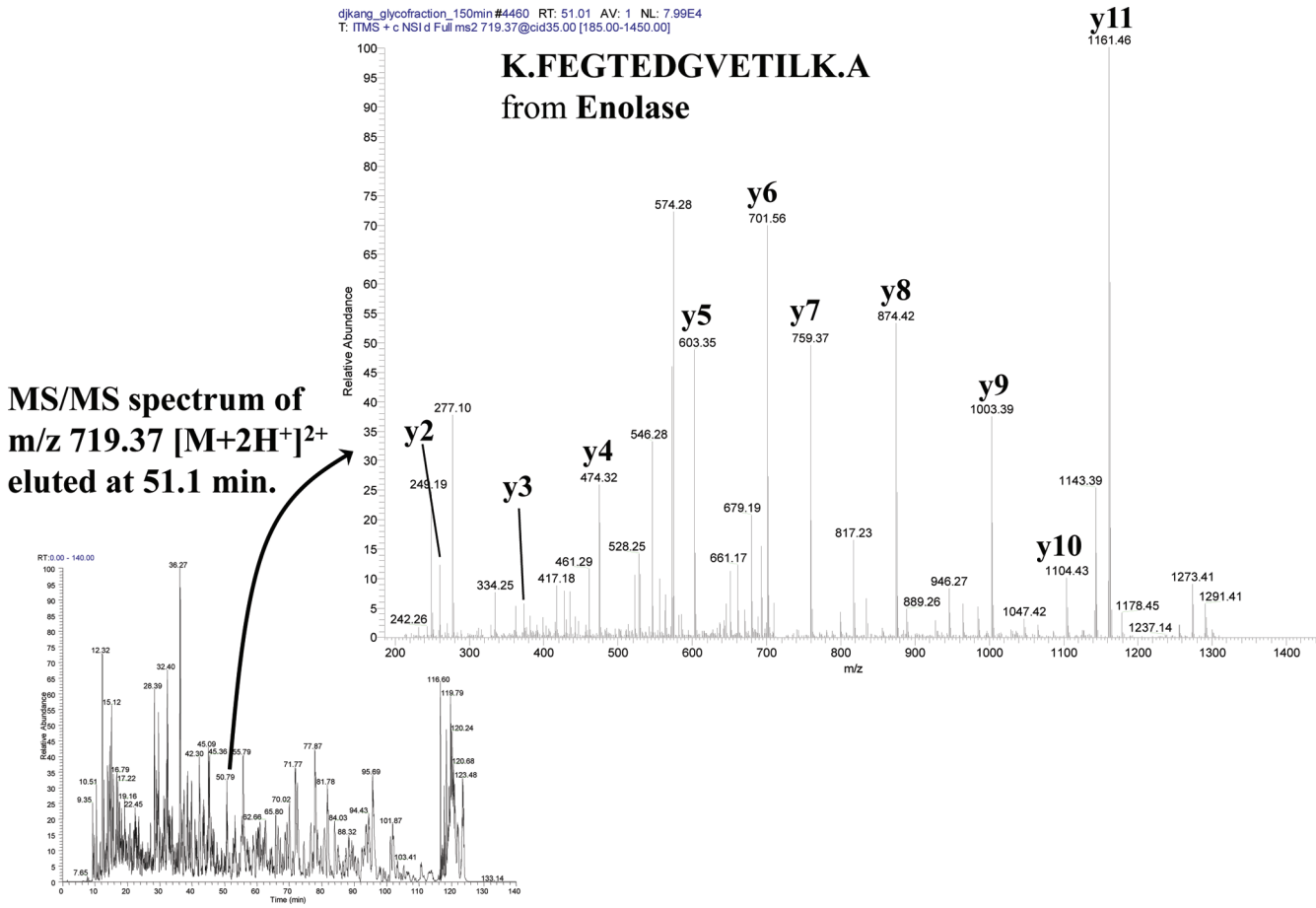


Figure 6. Precursor scan MS spectra at 51.1 min during RPLC-ESI-MS-MS run and the corresponding CID spectra of a precursor ion m/z 719.37 $[M + 2H^+]^{2+}$ were identified as a peptide of K.FEGTEDGVETILK.A from *Enolase* (RPLC-ESI-MS-MS run was obtained after the salt step elution with 2 mM NH_4HCO_3)

Table 1. Identification of ConA-Specific Binding Proteins and Their Cellular Functions for Collected ConA−Glycoproteins Fraction Using HF5

access number	identified proteins	M_w (kDa)	no. of peptides matched	functions	no. of N-glycosylation site(s)[a].
Q1JHQ6	enolase	47.3	16	glycolysis	1
Q8K872	elongation factor Tu	43.7	14	protein biosynthesis	1
Q5XBV6	glucosamine−fructose-6-phosphate aminotransferase	65.2	10	carbohydrate biosynthesis	1
P69883	60 kDa chaperonin	57.1	14	stress response	2
P60811	foldase protein prsA 1	38.5	7	protein folding	1
P08089	M protein, serotype 6	53.4	7	phagocytosis	8
Q5XE03	glycerol-3-phosphate dehydrogenase [NAD(P)+]	36.6	4	phospholipid biosynthesis	1
P0C0G7	glyceraldehyde-3-phosphate dehydrogenase	35.9	10	glycolysis	2
Q48RR3	chaperone protein dnaK	64.9	12	stress response	3
Q8K5Q0	**C5a peptidase**	**128.1**	6	virulence	7
P69887	triosephosphate isomerase	26.6	5	carbohydrate biosynthesis	-
Q5XAQ3	cysteine synthase	33.2	4	amino-acid biosynthesis	1
P65243	ribose-phosphate pyrophosphokinase 1	35.0	5	nucleotide biosynthesis	1
Q99ZE5	stress response regulator gls24 homologue	19.9	4	stress response	1
Q5XC60	probable NADH oxidase	49.5	3	electron transfer	3
Q99YG1	**translation initiation factor IF-2**	**105.4**	3	protein biosynthesis	6
P0C0B5	chaperone protein dnaJ	40.4	2	stress response	4
Q99ZD0	6-phosphofructokinase	35.7	2	glycolysis	1
Q8P137	Ribose-phosphate pyrophosphokinase 2	35.6	2	nucleotide biosynthesis	1
P68905	Fructose-bisphosphate aldolase	31.1	3	glycolysis	1
P60812	Foldase protein prsA 2	34.3	2	protein folding	4
Q8K7R7	Phosphoglucosamine mutase	48.3	3	carbohydrate biosynthesis	-
Q5XAJ6	Glycyl-tRNA synthetase beta subunit	75.1	12	protein biosynthesis	3
Q5XCL7	Probable ATP-dependent Clp protease ATP-binding subunit	77.4	5	electron transfer	4
P0A349	GTP-sensing transcriptional pleiotropic repressor codY	28.6	4	Transcription regulation	1
Q1JK21	Trigger factor	50.4	3	protein biosynthesis	3
P0C0I3	Streptolysin O	36.5	4	cytolysis	6
Q9A0K9	Septation ring formation regulator ezrA	66.0	3	cell division	4
P65757	manganese-dependent inorganic pyrophosphatase	33.5	4	electron transfer	1
Q1JCL2	ATP synthase epsilon chain	15.4	2	electron transfer	1
P0A4D7	Probable ATP-dependent RNA helicase exp9	58.7	4	electron transfer	2
P65666	Putative 2-dehydropantoate 2-reductase	33.8	2	carbohydrate biosynthesis	3
Q8K8P9	Glutamyl-tRNA synthetase	55.0	2	protein biosynthesis	1
P0C0H6	Inosine-5′-monophosphate dehydrogenase	52.7	2	carbohydrate biosynthesis	4
Q9A115	Threonyl-tRNA synthetase	74.2	2	protein biosynthesis	1
Q8K880	Lysyl-tRNA synthetase	56.5	4	protein biosynthesis	3
Q99XI2	Cobalt import ATP-binding protein cbiO 1	30.8	2	electron transfer	1
Q5XCG7	Probable dTDP-4-dehydrorhamnose 3,5-epimerase	22.4	2	carbohydrate biosynthesis	1
Q5XCM0	ATP-dependent Clp protease ATP-binding subunit clpX	44.9	3	protein folding	2
Q5XE96	**DNA-directed RNA polymerase subunit beta**	**135.2**	7	nucleotide biosynthesis	3
Q5XAC7	Glutamyl-tRNA(Gln) amidotransferase subunit A	52.2	2	protein biosynthesis	1
Q3JZM2	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	53.7	2	protein biosynthesis	1
Q8K5Z0	**UvrABC system protein A**	**103.9**	3	DNA repair	2
Q48SU7	S-adenosylmethionine synthetase	43.1	2	electron transfer	1
Q99ZA6	D-alanine--poly(phosphoribitol) ligase subunit 1	56.9	2	nucleotide biosynthesis	1

[a] A number of N-linked glycosylation site(s) of identified proteins were theoretically obtained by using a NetNGlyc 1.0 Server algorithm (www.cbs.dtu.dk/services/NetNGlyc).

From 2D SCX-RPLC-ESI-MS-MS analysis of the enriched ConA−glycoproteins fraction via HF5 run, 45 proteins were identified and listed in Table 1. Especially, most of the identified proteins, bound to the ConA−glycoproteins fraction, are less than 77.4 kDa in MW, and consequently, it can be expected that they are referred to as ConA-specific binding proteins, because a protein having a MW of ~77.4 kDa should be eluted within ~6 min (see Figure 2a). This means that identified proteins having an affinity for ConA are candidates for high mannose N-linked glycoproteins, except the four proteins such as C5a peptidase (128.1 kDa), translation initiation factor IF-2 (105.4 kDa), DNA-directed RNA polymerase (135.2 kDa), and UvrABC system protein A (103.9 kDa). The four proteins may elute along with ConA−glycoproteins complex, since their MWs are close to multimeric forms of ConA. Furthermore, all of the identified proteins listed in Table 1 were searched using a NetNGlyc 1.0 Server algorithm (www.cbs.dtu.dk/services/NetNGlyc) to verify the N-linked glycosylation. The investigation of N-linked glycosylation is generally carried out by two factors: (i) N-linked glycosylation has a well-defined consensus sequence as Asn-Xaa-Ser/Thr in which Xaa denotes any amino acid except proline; (ii) as a endoglycosidase, peptide *N*-glycosidase F (PNGase F) provides the determination of glycosylation site(s) among amino acids due to which this enzyme cleaves amino acids between the carbohydrate structure and asparagine; the asparagine residue is converted to aspartate, inducing a mass shift of +0.98. Unfortunately, enzymatic based-mass spectrometric analyses were not carried out in this study to determine those site(s) using a PNGase F. However, the former factor that is examined with consensus sequence of Asn-Xaa-Ser/Thr can be theoretically estimated by which N-linked glycosylation site(s) among identified proteins are confirmed

using a NetNGlyc 1.0 Server algorithm. Consequentially, 43 proteins, which were larger than the potential threshold of 0.5, have more than one consensus sequence out of total identified proteins, except triosephosphate isomerase and phosphoglucosamine mutase as listed in Table 1. It is evident that these proteins enriched by HF5 are ConA-specific binding proteins as N-linked glycosylated proteins. Additionally, on the basis of the MS analyses, several oxonium ions (e.g., m/z 186.077 of HexNAc-H20, m/z 204 of HexNAc, m/z 292.094 of NeuAc, m/z 295.094 of Pent(Hex), m/z of 366.132 of Hex(HexNAc), m/z 528.185 of Hex2(HexNAc), m/z 657.227 of NeuAc(Hex)(HexNAc), and so on) were observed during 2D SCX-RPLC-ESI-MS-MS analysis. However, we were not concerned with the fragment ions generated via the CID process for glycosylated peptides in this study, because there is not yet a proper database algorithm for identifying glycoproteins and their glycan structures from glycoprotein mixtures. Additionally, the fragment ions formed by CID of 7T LTQ-FTICR-MS was not suitable for the determination of the exact glycan structures of glycosylated peptides due to the difficulty of manual calculation for the charge states of each fragment ion, the limited mass range of ion optics ($\sim m/z$ 2000 in normal mode), and the suppression of ionization efficiency by coeluted glycopeptides and nonglycosylated ones. In this study, the HF5 has been evaluated as a molecular sorting method to provide an alternative tool to selectively isolate high mannose type N-linked glycoproteins from bacterial lysates, based on affinity for ConA.

Conclusion

In this study, we examined the possibility of enriching N-linked glycoproteins from bacterial lysates of *S. pyogenes* by using an affinity for ConA and molecular mass sorting by HF5. On the basis of 2D SCX-RPLC-ESI-MS-MS analysis, we identified 45 N-linked glycoproteins, which exist in more than one specific consensus-sequence Asn-Xaa-Ser/Thr where X can be any amino acid except proline. Consequently, it is confirmed that 45 identified proteins, enriched by using HF5, are candidates for high mannose type N-linked glycoproteins. From the perspective of a preparative method, HF5 can simultaneously fractionate the N-linked glycoproteins from the bacterial lysates without replying on conventional ConA immobilized methods in terms of a limited amount of glycoproteins enriched and low sample recovery via flow-through study. In addition, the total enriched volume of the ConA fraction that can be reduced to about 160 μL (40 μL/min of outflow rate $\times$ 4 min of HF5 time interval of collection) is more suitable for proteolytic digestion without an additional concentration step. A limitation of this study was the repeated runs (10 μL $\times$ 10 runs) due to the use of a microbore HF (450 μm-i.d.) in order to collect a proper amount of N-linked glycoproteins for secondary MS analysis. However, it can be overcome by employing a larger diameter hollow fiber ($\sim$ 1 mm-i.d.) or conventional FlFFF channel systems that can be injected with larger amount of sample materials. While the current study does not provide a complete characterization of the glycan structures of identified glycoproteins from the MS/MS spectra, it is promising that HF5 has a possibility as a new alternative method to enrich high mannose type N-linked glycoproteins from bacterial lysates.

Acknowledgment. This research was supported by the Converging Research Center Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2009-0081904). The authors also thank to Y.H.K. for the access to 7T LTQ-FTICR-MS.

References

(1) Tsur, D.; Tanner, S.; Zandi, E.; Bafna, V.; Pevzner, P. A. *Nat. Biotechnol.* **2005**, *23*, 1562–1567.
(2) Shu, H.; Chen, S.; Bi, Q.; Mumby, M.; Brekken, D. L. *Mol. Cell. Proteomics* **2004**, *3*, 279–286.
(3) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21*, 255–261.
(4) Madera, M.; Mechref, Y.; Klouckova, I.; Novotny, M. V. *J. Proteome Res.* **2006**, *5*, 2348–2363.
(5) Lowe, J. B.; Marth, J. D. *Annu. Rev. Biochem.* **2003**, *72*, 643–691.
(6) Wollscheid, B.; Bausch-Fluck, D.; Henderson, C.; O'Brien, R.; Bibel, M.; Schiess, R.; Aebersold, B.; Watts, J. D. *Nat. Biotechnol.* **2009**, *27*, 378–386.
(7) Kuster, B.; Krogh, T. N.; Mortz, E.; Harvey, D. J. *J. Proteome Res.* **2001**, *1*, 350–361.
(8) Alvarez-Manilla, G.; Atwood III, J.; Guo, Y.; Warren, N. L.; Orlando, R.; Pierce, M. *J. Proteome Res.* **2006**, *5*, 701–708.
(9) Jebanathirajah, J.; Steen, H.; Roepstorff, P. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (7), 777–784.
(10) Madera, M.; Mechref, Y.; Klouckova, I.; Novotny, M. V. *J. Proteome Res.* **2006**, *5* (9), 2348–2363.
(11) Madera, M.; Mechref, Y.; Novotny, M. V. *Anal. Chem.* **2009**, *77* (13), 4081–4090.
(12) Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T. *Nat. Biotechnol.* **2003**, *21*, 667–672.
(13) Bedair, M.; Oleschuk, R. D. *Analyst* **2006**, *131*, 1316–1321.
(14) Fournier, T.; Medjoubi-N, N.; Porquet, D. *Biochim. Biophys. Acta* **2000**, *1482*, 157–171.
(15) Treuheit, M. J.; Costello, C. E.; Halsall, H. B. *Biochem. J.* **1992**, *283*, 105–112.
(16) Giddings, J. C. *Anal. Chem.* **1981**, *53*, 1170A–1175A.
(17) Reschiglian, P.; Zattoni, A.; Roda, B.; Cinque, L.; Melucci, D.; Min, B. R.; Moon, M. H. *J. Chromatogr., A* **2003**, *985*, 519–529.
(18) Kang, D.; Moon, M. H. *Anal. Chem.* **2005**, *77*, 4207–4212.
(19) Giddings, J. C. *Science* **1993**, *260*, 1456–1465.
(20) Reschiglian, P.; Moon, M. H. *J. Proteomics* **2008**, *73*, 265–274.
(21) Reschiglian, P.; Zattoni, A.; Cinque, L.; Roda, B.; Dal-Piaz, F.; Roda, A.; Moon, M. H.; Min, B. R. *Anal. Chem.* **2004**, *76*, 2103–2111.
(22) Roda, A.; Parisi, D.; Guardigli, M.; Zattoni, A.; Reschiglian, P. *Anal. Chem.* **2006**, *78*, 1085–1092.
(23) Zattoni, A.; Rambaldi, D. C.; Roda, B.; Parisi, D.; Roda, A.; Moon, M. H; Reschiglian, P. *J. Chromatogr., A.* **2008**, *1183*, 135–142.
(24) Reschiglian, P.; Zattoni, A.; Roda, B.; Cinque, L.; Parisi, D.; Roda, A.; Dal-Piaz, F.; Moon, M. H.; Min, B. R. *Anal. Chem.* **2005**, *77*, 47–56.
(25) Kim, K. H.; Moon, M. H. *Anal. Chem.* **2009**, *81* (4), 1715–1721.
(26) Kang, D; Moon, M. H. *Anal. Chem.* **2006**, *78*, 5789–5798.
(27) Kang, D.; Oh, S.; Reschiglian, P.; Moon, M. H. *Analyst* **2008**, *133*, 505–515.
(28) Kang, D.; Oh, S.; Ahn, S.-M.; Lee, B.-H; Moon, M. H. *J. Proteome Res.* **2008**, *7* (8), 3475–3480.
(29) Kang, D.; Yoo, J. S.; Kim, M. O.; Moon, M. H. *J. Proteome Res.* **2009**, *8* (2), 982–991.
(30) Kim, K. H.; Kang, D.; Koo, H.; Moon, M. H. *J. Proteomics* **2008**, *71*, 123–131.
(31) Kang, D.; Nam, H.; Kim, Y. -S.; Moon, M. H. *J. Chromatogr., A* **2005**, *1070*, 193–200.
(32) Bang, D. Y.; Ahn, E.; Moon, M. H. *J. Chromatogr., B* **2007**, *852*, 268–277.
(33) Beres, S. B.; Richter, E. W.; Nagiec, M. J.; Sumby, P.; Porcella, S. F.; DeLeo, F. R.; Musser, J. M. *Proc. Natl. Acad. Sci. U.S.A.* , *103*, 7059–7064.
(34) Green, N. M.; Zhang, S.; Porcella, S. F.; Nagiec, M. J.; Barbian, K. D.; Beres, S. B.; LeFebvre, R. B.; Musser, J. M. *J. Infect. Dis.* **2005**, *192* (5), 760–770.

PR900937W