analytical chemistry

Article

pubs.acs.org/ac

Dual Lectin-Based Size Sorting Strategy to Enrich Targeted N-Glycopeptides by Asymmetrical Flow Field-Flow Fractionation: Profiling Lung Cancer Biomarkers

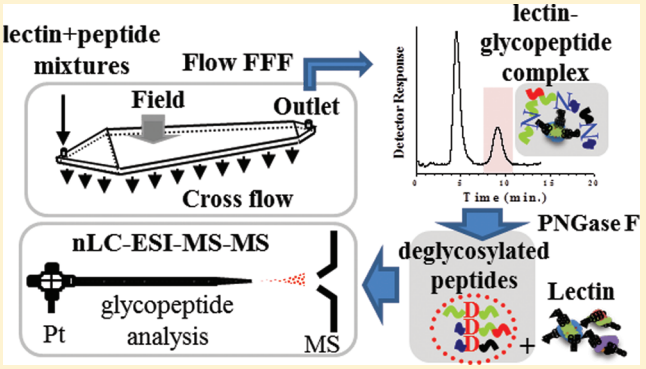
Jin Yong Kim,[†] Sook-Kyung Kim,[‡] Dukjin Kang,*[,‡] and Myeong Hee Moon*[,†]

[†]Department of Chemistry, Yonsei University, Seoul, 120-749, Korea

[‡]Center for Bioanalysis, Division of Metrology for Quality of Life, Korea Research Institute of Standards and Science, Daejeon, 305-340, Korea

Supporting Information

ABSTRACT: A dual lectin-based size sorting and simultaneous enrichment strategy for selectively isolating N-linked glycopeptides was developed using asymmetrical flow field-flow fractionation (AF4). AF4 is an elution-based method for separating biological macromolecules that has been utilized for the separation of lectin−glycopeptide complexes formed by mixing serum peptides with lectin cocktails according to the difference in diffusion coefficients. It has also been used for simultaneous depletion of nonglycosylated peptides. The dual lectin-based enrichment method was applied to proteolytic peptides from lung cancer serum samples with two lectins (WGA, GlcNAc-specific, and SNA, Sia-specific), and the whole mixture was separated by AF4. The lectin−glycopeptide complex fractions collected during AF4 separation were endoglycosidically digested with PNGase F. The resulting deamidated glycopeptides were analyzed by nanoflow liquid chromatography-electrospray ionization-tandem mass spectrometry (nLC-ESI-MS-MS) to semiquantitatively profile the N-linked glycopeptides from the sera of lung cancer patients and healthy controls. The AF4 enrichment strategy coupled with nLC-ESI-MS-MS identified 16/24 (up/down-regulated by at least 10-fold compared to normal sera) N-linked glycopeptides from a WGA complex fraction of lung cancer sera and 18/3 from a SNA fraction.



Glycosylation is a common post-translational modification (PTM) of proteins in eukaryotes. It has been widely studied as a clinical route for the early detection of human diseases and for the elucidation of biological functions related to cell proliferation.[1−4] With N-linked glycosylation, micro-heterogeneous glycans are attached through asparagine site(s) of proteins. N-linked glycosylation has a well-defined consensus sequence of Asn-Xaa-Ser/Thr (Xaa denotes any amino acid except proline) that induces changes in the biochemical process such as cell−cell interaction, maintenance of cell structure, cell signaling, and protein folding.[2,5−8] Despite the important roles of glycosylation, it has not yet been thoroughly characterized. A major difficulty in studying N-glycosylation lies in the complicated purification steps required to isolate targeted N-linked glycoprotein from biological samples (e.g., biological fluids, tissue, and vesicles). It is desirable to develop a simple and highly efficient isolation method for characterizing biological alterations related to human diseases on the basis of qualitative/quantitative proteomic analysis.[8−10]

Recently, the application of advanced tandem mass spectrometry (MS-MS)-based strategies coupled with on-/off-line multidimensional separation techniques for glycoproteins has been used to rapidly elucidate the roles of targeted N-linked glycoproteins in cellular processes.[11−13] Among advanced MS-MS-based techniques, electron capture dissociation (ECD)/electron-transfer dissociation (ETD) reactions have been employed as complementary strategies for identifying N-linked glycopeptides and their glycan structures. This can occur without endoglycosidic digestion, since ECD/ETD reactions of N-linked glycopeptides can preferentially cleave the peptide backbone while keeping their glycan structures intact.[9,14−16] The structural information obtained by MS-MS experiments on N-linked glycopeptides provides a high confidence level in the identification of N-linked glycoproteins and their glycan structures. Despite advanced tandem MS-based approaches, it is difficult to completely characterize an N-linked glycoprotein out of complicated protein mixtures since proteolytically digested glycopeptides are extremely infrequent, comprising only about 2−5% of the total amount of peptides. Moreover, the ionization efficiency of glycopeptides is relatively poorer than that of ordinary peptides. For this reason, coelution with a

Received: March 19, 2012
Accepted: May 19, 2012
Published: May 20, 2012

 © 2012 American Chemical Society dx.doi.org/10.1021/ac300772w | Anal. Chem. 2012, 84, 5343−5350

large excess of peptides may suppress the ionization of targeted glycopeptides.[17,18]

To overcome the bottlenecks mentioned above, enrichment methods for N-linked glycoproteins prior to advanced MS-MS experiments have recently been advanced by lectin-based affinity chromatography (LAC),[11,12,19,20] hydrophilic interaction chromatography (HILIC),[5] graphitic column chromatography (GC),[21] size exclusion chromatography (SEC),[17] and hollow fiber flow field-flow fractionation (HF5).[22] Among these methods, the lectin-based approach has been widely utilized to isolate targeted N-linked glycopeptides/glycoproteins selectively from peptide/protein mixtures in combination with different immobilization platforms such as monolithic columns, protein chips, membranes, and nanosized beads.[2,9,23] Lectins are particularly capable of binding with microheterogeneous glycans such as mannose (Man), *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), sialic acid (Sia), and fucose (Fuc).[9,20,24] Lectin-based approaches are promising for three key reasons: selective isolation of targeted glycoproteins/glycopeptides from protein/peptide mixtures, the capability to infer compositional information about glycan structures according to lectin types, and compatibility with size sorting methods such as HF5, SEC, and gel electrophoresis (GE)[9] in size differentiation of targeted glycoproteome via lectin−glycan interaction.

Asymmetrical flow field-flow fractionation (AFlFFF or AF4) is an elution-based alternative tool for proteomics that has been used for separating biological components such as proteins, DNA, subcellular organelles, biological vesicles, and cells.[22,25−33] Unlike a gel-based method, protein separation in AF4 can be achieved without denaturation. Because of this, AF4 is a good alternative for characterizing intact proteomic features such as protein complexation and antigen−antibody interaction. Moreover, nongel-based two-dimensional separation by online isoelectric focusing (IEF) with AF4 or HF5 provides the potential to overcome weak points (e.g., labor-intensive, time-consuming, protein denaturation) of conventional gel-based approaches.[29,31] The IEF-AF4 technique was applied to fractionate intact and low abundance proteins from the human urinary proteome according to their isoelectric point (pI) and hydrodynamic diameter (d_s). Very recently, a chip-type AF4 was directly coupled with MS and applied for top-down intact protein characterization.[34] Protein size sorting with AF4 is achieved by applying cross-flow in the direction perpendicular to the migration flow through a flat membrane of an AF4 channel wall.[28,32,34] Briefly, the separation of proteins (or lectins) can be achieved in an empty channel space with a rectangular cross-section. When proteins are introduced to an inlet port of the AF4 channel, they are differentially distributed from the accumulation wall of the AF4 channel. They reach equilibrium positions that are dependent on the two opposing forces of cross-flow as an external field and diffusion of sample components. Due to differences in the diffusion coefficients of proteins, smaller proteins will be elevated further away from the accumulation wall and entrained in the faster laminar flow streamline, leading to an earlier elution than larger proteins. Therefore, protein size sorting by AF4 is achieved in increasing order of diffusion coefficients or hydrodynamic diameters.

In a previous report, we demonstrated that HF5 can be utilized to enrich N-linked glycoproteins from bacterial lysates by combining the specific binding affinity of Concanavalin A (ConA) to high mannose-type glycans of glycoproteins with size sorting capability.[22] The HF5-based enrichment method allowed glycoprotein isolation at the protein level. However, the selective isolation of targeted N-linked glycoproteins from protein mixtures by lectin species can be hindered when a protein glycosylation site is blocked by the three-dimensional structure of proteins. In addition, multiple HF5 runs were required to accumulate N-linked glycoproteins.

In this study, an experimental strategy was developed to utilize AF4 as an enrichment device for sorting N-linked glycopeptides from proteolytic peptide mixtures by the difference in hydrodynamic diameters of lectin−glycopeptide complexes and the simultaneous removal of unbound nonglycosylated peptides. The selective recognition of glycopeptides to each lectin was evaluated with a lectin mixture (WGA and SNA). This was followed by AF4 separation of a lectin−glycopeptide complex. The developed principle can reduce the drawbacks of glycoprotein enrichment in protein level, since some large unbound proteins are eluted together with the lectin−glycoprotein complex, thereby lowering the isolation efficiency of glycoproteins. Glycopeptide enrichment with AF4 can remove nonglycosylated peptides through cross-flow during the focusing/relaxation of AF4 operation, and it will result in the elution of only lectin−glycopeptide complexes via size separation. The proposed AF4-based enrichment method was initially evaluated using α-1-acid glycoprotein (AGP) with five N-linked glycosylation sites and its affinity to ConA.[35] The lectin−glycopeptide fraction collected during AF4 separation was endoglycosidically digested using PNGase F and analyzed by nanoflow liquid chromatography-electrospray ionization-tandem mass spectrometry (nLC-ESI-MS-MS) to identify N-linked glycopeptides and their glycosylation site(s).[22,36] The developed method was applied to proteolytic peptides of human sera from lung cancer patients and healthy controls by utilizing a lectin mixture of WGA (GlcNAc-specific) and SNA (Sia-specific)[20] for the simultaneous lectin-specific isolation of glycopeptides by AF4. Both AF4 fractions (WGA−glycopeptide and SNA−glycopeptide complexes) were treated with PNGaseF, and the resulting deamidated peptides were analyzed by nLC-ESI-MS-MS to quantitatively profile N-linked glycopeptides from the sera of lung cancer patients and healthy controls.

MATERIALS AND METHODS

Materials. α-1-Acid glycoprotein (AGP), concanavalin A (ConA), lectin from *Triticum vulgaris* (WGA, ∼18 kDa/subunit), lectin from *Sambucus nigra* (SNA, ∼38 kDa/subunit), and the five protein standards were purchased from Sigma (St. Louis, MO, USA). The protein standards were carbonic anhydrase (CA, 29 kDa), bovine serum albumin (BSA, 66 kDa), alcohol dehydrogenase (ADH, 150 kDa), apoferritin (444 kDa), and thyroglobulin (669 kDa).

Serum samples were obtained from three patients with lung cancer and from three healthy normal persons at Severance Hospital (Seoul, Korea) after obtaining written consent. The ProteaPrep Albumin and IgG Depletion Sample Prep Kit from Protea Biosciences, Inc. (WV, USA) was utilized to deplete abundant proteins from the serum samples. After depletion, the solvent in the serum eluate was replaced with binding buffer solution (1 mM $CaCl_2$, 1 mM $MgCl_2$, and 1 mM $MnCl_2$ in 50 mM TBS). The serum protein concentration of the final eluate was measured by the Bradford assay.

Silica capillary tubes used to prepare capillary LC columns (75 μm i.d.) and plumbing capillary LC (20, 75, and 100 μm i.d.) were purchased from Polymicro Technology LLC

dx.doi.org/10.1021/ac300772w | *Anal. Chem.* 2012, 84, 5343−5350

(Phoenix, AZ, USA). All had the same outer diameter (OD) of 360 μm.

Proteolytic Digestion and Lectin−Glycopeptide Complexation. See Supporting Information.

Asymmetrical Flow Field-Flow Fractionation (AF4). The model Eclipse3 AF4 channel from Wyatt Technology Europe GmbH (Dernbach, Germany) utilized to size sort lectin−glycopeptide complexes had the following channel dimensions: tip-to-tip length, 26.6 cm; thickness, 250 μm; and trapezoidal breadth decreasing from 2.2 to 0.6 cm to the outlet. At the accumulation wall of the AF4 channel, a Nadir cellulose membrane (MWCO 10 kDa) supplied from Wyatt Technology Europe GmbH was used to prevent sample penetration by cross-flow. During the separation of lectin−glycopeptide complexes, free unbound peptides from the sample mixture were depleted through the channel membrane along the cross-flow exit. A model SP930D HPLC pump from Young-Lin Instruments (Seoul, Korea) was used to deliver the sample and carrier liquid to the AF4 channel. During injection and focusing/relaxation, the pump flow was divided into two parts leading to the channel inlet and outlet at a ratio of 1:9, allowing for 1.5 min including sample injection and focusing/relaxation period. Detection was performed at 280 nm with a model M720 UV detector, and the detector signals were recorded using Autochro-Win 2.0 plus software, all from Young-Lin. The AF4 carrier liquid was the same as the buffer solution used for lectin−glycopeptide complexation. It was prepared with ultrapure (>18 MΩ·cm) water and filtered with a nitrocellulose membrane filter (0.22 μm) from Millipore (Danvers, MA, USA) prior to use.

Endoglycosidic Digestion of N-Linked Glycopeptides Bound to Lectins. See Supporting Information.

nLC-ESI-MS-MS. Nanoflow LC-ESI-MS-MS experiments were carried out with a LTQ Velos ion trap mass spectrometer from Thermo Finnigan (San Jose, CA, USA) interfaced with a model 1260 capillary LC system from Agilent Technologies (Waldbronn, Germany). Prior to the analytical capillary pulled tip column [17 cm × 75 μm i.d., 5 μm-100 Å Magic C18AQ from Michrom Bioresources Inc. (Auburn, CA, USA)], a trap column (3 cm × 200 μm i.d., 5 μm-200 Å Magic C18AQ) was connected for sample desalting via a PEEK microcross from Upchurch Scientific. Columns were packed in the laboratory following detailed procedures reported in the literature.[22,29] The mobile phase composition of the binary gradient elution for nLC was 98/2 (v/v) water/acetonitrile for mobile phase A and 95/5 acetonitrile/water for B, and both were added with 0.1% formic acid. Deamidated peptide sample was loaded into the trap column first by 2% B for 10 min, and then, gradient elution began with an initial increase to 10% B for 1 min. After a gradual increase from 10% B to 30% for 59 min, the column was cleaned by ramping up to 80% B for 3 min, maintaining this for 10 min, and then returning to 2% for 2 min. For at least 15 min, the column was allowed to re-equilibrate.

For ESI, the column outflow rate was kept at 200 nL/min by controlling the length of the capillary tube (20 μm i.d) attached to the microcross. The ESI voltage was 2.5 kV for the MS precursor scan (m/z 300−1800). Data-dependent MS-MS scans were carried out for three prominent precursor ions from each precursor scan. MS-MS spectra were analyzed using Proteome Discoverer software (Ver 1.2.0.208) from Thermo-Finnigan with a human database from nrNCBI. The mass tolerance between the measured and calculated mass was set as 1.0 Da for the molar mass of a precursor peptide and 0.8 Da for that of fragment ions. For the validity screening, only proteins searched with the following requirements were selected: ΔCn scores of 0.1 and cross-correlation (Xcorr) values larger than 2.4, 2.8, and 3.7 for singly-, doubly-, and triply charged ions, respectively. The fixed modification was set as carbamidomethylation of cysteine along with oxidation of methionine and deamidation of asparagine. When the heterogeneous glycopeptides, overlapped with the same glycosylation site(s) due to the steric hindrance of proteolysis by glycans, were identified, only the peptide having the highest score was selected to enhance the confidence level in the assignment of the glycosylation site and to keep the reproducibility in quantification between lung cancer and control sera.

RESULTS AND DISCUSSION

The utility of AF4 as a dual lectin-based size sorting and glycopeptide enrichment method was evaluated with tryptic peptides of AGP by fractionating a target lectin−glycopeptide complex. Prior to glycopeptide enrichment, the size sorting capability of AF4 was confirmed by the separation of five protein standards (29−669 kDa) (Figure 1a) obtained at $\dot{V}_{out}$/

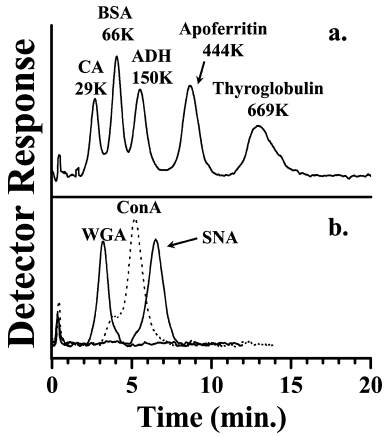


Figure 1. (a) AF4 separation of five protein standards (CA: 29 kDa, BSA: 66 kDa, ADH: 150 kDa, apoferritin: 444 kDa, and thyroglobulin: 669 kDa) with injection amounts of 2 μg for each species, and (b) the superimposed AF4 fractograms of lectins (WGA, ConA, and SNA) with individual injection. Flow rate conditions were an outflow rate of 0.5 mL/min and a cross-flow rate of 4.5 mL/min.

$\dot{V}_c$ (outflow/crossflow rates) = 0.5/4.5 in mL/min. Due to the capability of AF4 in handling intact proteins, each of the three protein standards (ADH, apoferritin, and thyroglobulin) had its own assembly of multimeric subunits eluted without dissociation. Under the same run conditions, three lectins (WGA, ConA, and SNA) were injected individually into the AF4 channel as shown in Figure 1b in order to estimate apparent sizes in the AF4 carrier solution. While each subunit of these lectin molecules has a small molar mass (~18 kDa/subunit for WGA, ~26 kDa/subunit for ConA, and ~38 kDa/subunit for SNA), the three lectins eluted at retention times corresponding to much larger molecules (compared to Figure 1a). This result supported the fact that the lectins eluted with multimeric conformations: WGA as dimers, ConA as tetramers to hexamers, and SNA as octamers. The multimeric form of lectin can be spontaneously generated depending on the pH of the surrounding aqueous solution, as was confirmed by our previous study.[22] The formation of multimers of each lectin facilitates homogeneous lectin−glycan interactions and their

dx.doi.org/10.1021/ac300772w | Anal. Chem. 2012, 84, 5343−5350

isolation from other nonspecific binding caused by impure species in lectins since lectin multimer−glycan complexes become much larger and can be separated by sizes from impure species in AF4. From Figure 1b, WGA and SNA were selected as multiplexed enrichment devices for targeted (GlcNAc- and Sia-specific) N-glycopeptides from peptide mixtures.

To examine the efficiency of selective glycopeptide enrichment, the tryptic peptides of α-1-acid glycoprotein (AGP), a typical glycoprotein standard, were utilized with ConA, which has specific affinity for high-mannose glycopeptides. AGP contains five N-linked glycosylation sites of N^{15}, N^{38}, N^{54}, N^{75}, and N^{85} and about 20% of mannose out of total glycans are post-translationally modified. ConA and the tryptic peptides of AGP were mixed at a ratio of 1:20 (total 10 μg). Then, 2 μg of ConA/peptide mixture was injected into the AF4 channel for size sorting. When the whole mixture was run through the AF4 channel, only the ConA−glycopeptide complex eventually eluted at the detector without unbound peptides, which were washed through the channel membrane (MWCO 10 kDa) by cross-flow (the fractogram is not shown). The ConA−glycopeptide fraction was collected during an AF4 run and treated with PNGase F for endoglycosidic digestion (or deamidation). Enzymatically, PNGase F provides evidence of the glycosylation site(s) of a glycopeptide sequence since it cleaves glycans between the innermost GlcNAc and asparagine of N-linked glycopeptides. As a result, the deamidation determines which asparagine residue is converted to aspartic acid with a mass shift of +0.98.

Figure 2 shows a comparison of two base peak chromatograms (BPCs) obtained by nLC-ESI-MS-MS of the tryptic peptide mixtures of AGP without size sorting (bottom) and of the endoglycosidically digested peptides (top) of the ConA−glycopeptide complex fraction collected during a single AF4 run. The two BPCs showed a dramatic difference in the number of eluted peptide peaks, which supported the fact that the majority of peptides were depleted during AF4 separation. The top part of Figure 2 shows two CID spectra of deamidated peptide ions (m/z of 958.9 $[M + 2H^+]^{2+}$ and m/z 961.5 $[M + 2H^+]^{2+}$) identified as QDQCIYn*TTYLNVQR (N^{75}) and SVQEIQATFFYFTPn*K (N^{54}), respectively, through a search of the nrNCBI database. After the database search of all detected peaks from the top part of Figure 2, all five N-linked glycopeptides containing the modification site as a result of asparagine deamindation were unambiguously detected out of all tryptic AGP peptides without detecting other non-glycosylated peptides.

This experiment demonstrated that using AF4 to sort lectin−glycopeptide complexes is promising for the selective isolation of N-linked glycopeptides without additional depletion steps. The depletion of ordinary peptides or glycosylated peptides with no affinity for specific lectins employed for an AF4 run can be achieved by coelution with a nonretained void peak at the beginning of separation. This can also be achieved by washing them through the channel membrane along with cross-flow exiting.

The developed strategy was applied to human serum samples to simultaneously enrich targeted N-linked glycopeptides using a mixture of lectins: WGA (GlcNAc-specific) and SNA (Sia-specific). Prior to the size sorting enrichment of serum glycopeptides, the serum sample was treated with an albumin/IgG depletion antibody cartridge to deplete albumin and IgG, the highly abundant protein components in human serum. After depleting albumin/IgG, the remaining serum

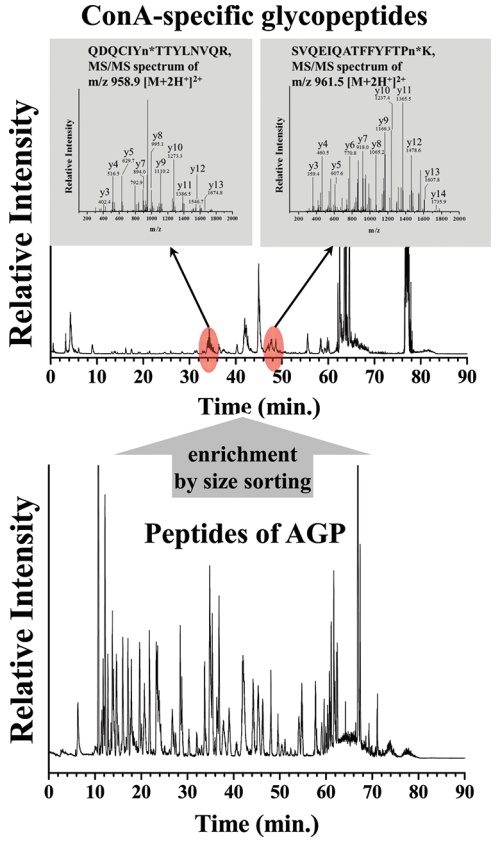


Figure 2. Base peak chromatograms (BPCs) of the tryptic peptide mixture of AGP (bottom) and of the endoglycosidically digested peptides (top) of the ConA−glycopeptide complex fraction collected during AF4. Insets are the data-dependent CID spectra of deamidated peptide ions identified as QDQCIYn*TTYLNVQR (N^{75}) for precursor ion m/z of 958.9 and SVQEIQATFFYFTPn*K (N^{54}) for m/z 961.5.

sample was digested in sequence using chymotrypsin and trypsin to enhance the efficiency of enzymatic cleavage for targeted amino acid site(s). This was done because the steric influence of N-linked glycosylation can increase the chances of proteolytic miscleavages. The latter effect had negative influences on the ESI ionization efficiency and on accurate identification of glycopeptides and their glycosylation sites from tandem MS spectra. After dual enzymatic digestion, 10 μg of serum peptides was mixed with 200 μg of WGA and 200 μg of SNA.

Figure 3a shows the AF4 fractogram of the lectin/peptide mixtures (5 μg injected) representing the elution of only lectin−glycopeptide complex peaks obtained under the same run conditions used in Figure 1. When the proteolytic peptides of human serum added to both lectins for 1 h at room temperature were injected to AF4, each lectin−glycopeptide complex eluted at an increased retention time (from 3.2 to 4.5 min for the WGA complex and from 6.5 to 9.1 min for the SNA complex in Figure 3a) compared to that of unconjugated lectin molecules, as shown in Figure 1b. It was not apparent whether the specific binding of lectin with glycopeptides induced an alteration in the apparent size of individual lectins by forming lectin-N-linked glycopeptide complexes. It is presumed that these increments originated from multimeric lectins binding with multiple numbers of N-linked glycopeptides (about 3−8 kDa each), with the exception of their miscleavages.[37]

dx.doi.org/10.1021/ac300772w | Anal. Chem. 2012, 84, 5343−5350

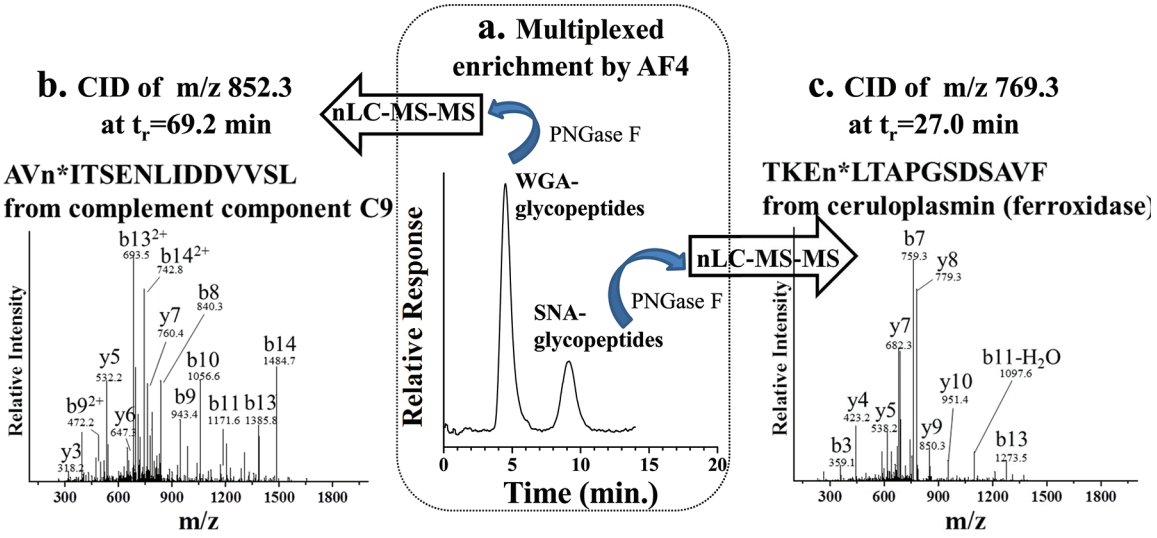


Figure 3. (a) AF4 separation of lectin/peptide mixtures that show only two peaks representing WGA- and SNA-specific glycopeptide complexes separated by size. Data-dependent CID spectra of precursor ions with (b) m/z 852.3 (t_r = 69.2 min) and (c) m/z 769.3 (t_r = 27.0 min) during each nLC-ESI-MS-MS run of deamidated glycopeptide mixtures endoglycosidically digested from the WGA–glycopeptide fraction and the SNA–glycopeptide fraction, respectively.

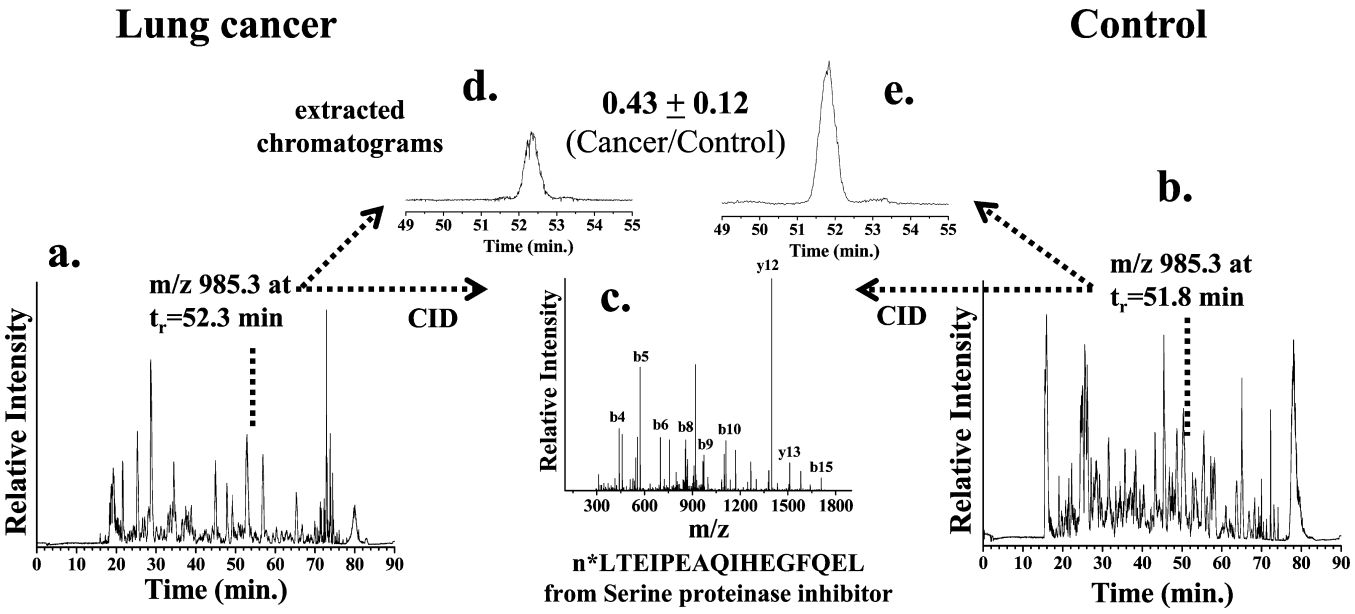


Figure 4. Comparison of BPCs of deamidated glycopeptide mixtures from each WGA–glycopeptide fraction of (a) lung cancer and (b) control sera obtained by nLC-ESI-MS-MS. Data-dependent CID spectra of each common glycopeptide from (c) WGA–glycopeptide fractions during nLC runs. Extracted ion chromatograms of precursor ions m/z 985.3 from nLC-ESI-MS runs of the WGA fraction from (d) lung cancer and (e) control samples.

Each lectin–glycopeptide complex fraction in Figure 3a was collected and endoglycosidically digested using PNGase F for nLC-ESI-MS-MS. Figure 3b shows the CID spectra of a precursor ion (m/z 852.3 3, $[M + 2H^+]^{2+}$) eluted at t_r = 69.2 min during nLC-ESI-MS-MS of the endoglycosidic peptide mixtures from the WGA–glycopeptide fraction. This resulted in a sequence identification of AVn*ITSENLIDDVVSL from complement component C9, which plays a key role in the innate and adaptive immune responses by forming pores in the plasma membrane of target cells. C9 has been used as a biomarker for various diseases (e.g., Huntington's disease, diabetes, and sepsis).[38,39] The CID spectra in Figure 3c are from the precursor ions (m/z 769.3, $[M + 2H^+]^{2+}$) eluted at 27.0 min during a separate nLC-ESI-MS-MS run from the SNA–glycopeptide fraction. The peptide sequence was identified as TKEn*LTAPGSDSAVF from ceruloplasmin, a well-known biomarker of Parkinson's disease.[40] nLC-ESI-MS-MS analysis of each fraction resulted in assigning 72 N-linked glycopeptides from the WGA–glycopeptides fraction and 39 from the SNA–glycopeptides fraction. These findings indicate that size separation of lectin–glycopeptide complexes by AF4 offers a new strategy for simultaneous multiplexed enrichment of targeted N-linked glycopeptides from peptide mixtures based on the intrinsic affinity for glycans.

AF4 with lectins was used as a size sorting strategy to quantitatively profile lectin-specific N-linked glycopeptides of

dx.doi.org/10.1021/ac300772w | Anal. Chem. 2012, 84, 5343–5350

Table 1. Up- or Down-Regulated (>10 times) N-Linked Glycoproteins Identified from Both WGA- and SNA-Bound Glycopeptide Fractions by AF4 Separation Followed by nLC-ESI-MS-MS

WGA bound	Up-regulated: Down syndrome cell adhesion molecule, Rev3-like catalytic subunit of DNA polymerase zeta, protein kinase (DNA-activated, catalytic polypeptide), Orosomucoid-1 (alpha-1-acid glycoprotein1), complement component 3 precursor, vitronectin precursor, complement component 4B proprotein, coagulation factor II precursor (prothrombin), complement factor H, serine/threonine protein kinase MAK, coagulation factor VIII isoform a precursor, arachidonate 15-lipoxygenase, Xaa-Pro dipeptidase, golgi autoantigen, transcription factor 1, Rho-GTPase-activating protein 7, protocadherin beta 1 precursor, tetraspan NET-7
	Down-regulated: beta-2-glycoprotein I precursor, transforming growth factor beta-stimulated protein TSC-22, histidine-proline rich glycoprotein, transcription elongation factor A (SII)-like 1, complement component C9, interalpha (globulin) inhibitor H2, hemopexin, alpha-2-macroglobulin precursor, ceruloplasmin (ferroxidase), transferrin, alpha 1B-glycoprotein, pregnancy-zone protein, plastin 3, lumican, complement component 1 inhibitor precursor, myeloperoxidase, elongin C (transcription elongation factor B), lipase H, F-box and leucine-rich repeat protein 2, complement factor B, vacuolar protein sorting 13D, carboxypeptidase E precursor, MAX dimerization protein 5, complement factor H
SNA bound	Up-regulated: complement factor B, complement component 4B proprotein, coagulation factor VIII isoform a precursor, coagulation factor II precursor (prothrombin), haptoglobin, hemopexin, serine/cysteine proteinase inhibitor (antitrypsin), ceruloplasmin (ferroxidase), Alpha-2HS-glycoprotein, ubiquitin protein ligase E3 component, C130007D14 protein, ataxia telangiectasia and Rad3 related protein, tigger transposable element derived 2, gamma-aminobutyric acid A receptor, SET and MYND domain containing 4, Lipase A, arachidonate 15-lipoxygenase, eukaryotic translation initiation factor 2
	Down-regulated: ubiquitin specific protease 8, dynein (heavy polypeptide 8), Alpha-2HS-glycoprotein

normal and lung cancer sera to unveil the determinants related to lung cancer. The same experimental procedures were applied for lung cancer and normal serum samples. Both WGA− and SNA−glycopeptide complex fractions were collected during AF4 separation of each digested peptide mixture (equivalent to 10 μg of serum proteins) with a lectin cocktail. To assess the relative differences between normal and cancer sera, three samples each from cancer and control sera were individually analyzed. The same precursor ions of m/z 985.3 $[M + 2H^+]^{2+}$ were detected at t_r = 52.3 (a lung cancer sample) and 51.8 min (a control sample) during a nLC-ESI-MS run of each endoglycosidic peptide mixture from each WGA−glycopeptide fraction, as shown in Figure 4a,b, respectively. These ions were identified as n*LTEIPEAQIHEGFQEL, which originated from the serine proteinase inhibitor (MS-MS spectra shown in Figure 4c). On the basis of a peak area comparison of the precursor ions from the extracted ion chromatogram shown in Figure 4d,e, the peak area ratio between lung cancer sera and controls was calculated to be 0.43 ± 0.12 (n = 3). The serine proteinase inhibitor (alpha-1-antitrypsin) is a well-known marker of early asbestosis diagnosis and plays a role in preventing lung inflammation caused by cigarette smoking, infection, and airborne particles. Consequently, other work has verified that a reduction of its quantity in serum simultaneously enhances the pathogenic potential of lung cancer.[41] Orosomucoid-1, a biomarker for an inflamed pilonidal abscess, was identified from both SNA−glycopeptide fractions. It showed a ratio of 0.47 ± 0.09 (cancer/control), as shown in Supplementary Figure 1, Supporting Information. Quantification was based on the precursor ions of m/z 1063.4 $[M + 2H^+]^{2+}$ with the identified peptide sequence of CANLVPVPITn*ATLDQITGK. The same ion was detected at t_r = 65.9 and 65.0 min for the nLC-ESI-MS-MS of each SNA fraction from lung cancer and control sera, respectively. Comprehensive examinations of N-linked glycopeptides bound to WGA revealed that the three different peptides were significantly increased in the lung cancer sample. SRFTKLn*ESTFDTQITK (10.08 ± 1.66 of cancer/control samples) from protein kinase, DNA-activated, catalytic polypeptide,[42] CANLVPVPITn*ATLDQITGK (11.89 ± 4.16) from orosomucoid-1,[43] TVLTPATNHMGn*VTFTIPANR (11.07 ± 1.02) from complement component 3,[44] and two peptides (TVKIPAMITSHPn*TTIAIK (0.11 ± 0.05) from down syndrome cell adhesion molecule like 1 and QKRQQEAQNAn*TTQDPL (0.12 ± 0.07) from the REV3-like and catalytic subunit of DNA polymerase zeta (REV3)) were found to be down-regulated for the cancer sample. These are included in up- or down-regulated (>10 times) glycoproteins compared to the control sample listed in Table 1, which is a summary from complete lists in Supplementary Table I (Supporting Information) of identified glycopeptides showing at least 2.5 times more relative changes.

Analysis of SNA-bound glycopeptides showed less significant alterations between cancer and control samples than analysis of WGA-bound glycoproteins. The exception was the same peptide, QKRQQEAQNAn*TTQDPL, which was increased 7.94 ± 3.02 times in the SNA fraction of the cancer sample while it was found to be 0.12 ± 0.07 (8.33-fold decrease) from the WGA fraction. The N-linked glycopeptide sequence of REV3 might be substantially promoted to form sialylated glycans instead of N-acetylglucosamidated ones in lung cancer sera. Interestingly, a chemotherapeutic study showed that reducing the REV3 level in a cell leads to significant extension in the survival rate of mice with lung tumors.[45] However, the N-linked glycosylation of REV3 in lung cancer has not yet been proven. Consequently, it is necessary to further examine which N-linked glycosylation of REV3 may be an alternative candidate for early lung cancer detection.

AF4 coupled with nLC-ESI-MS-MS resulted in the identification of 53 WGA-specific binding glycopeptides and 37 SNA-specific ones from lung cancer sera, and 59 WGA- and 25 SNA-specific glycopeptides from control samples as listed in Supplementary Table I (Supporting Information). Unfortunately, identified numbers of N-linked glycopeptides from both lung cancer and control sera in this study are less than those derived from a typical conventional lectin-immobilized column or cartridge. The low recovery of lectin-specific glycopeptides via AF4 size sorting may be caused by an intrinsic barrier such as weak lectin−glycan interactions and by a collection of lectin−glycopeptide fraction from a single AF4 run. However, this limitation can be overcome by employing antibodies with higher binding affinity with glycans or by accumulating more of each lectin−glycopeptide fraction from repeated AF4 runs or by lowering cross-flow rate condition to reduce field strength, in which the latter effect may reduce unexpected dissociation of glycopeptide from lectin. In addition, a series of endoglycosidic digestions by coupling PNGase F with PNGase A can be an alternative solution to increase the recovery of targeted glycopeptides since PNGase F is not universal to identify all N-linked glycopeptides. Aside from optimizing AF4 run conditions to increase the number of identified glycopeptides, we also performed a semiquantitative comparison of the peak areas of N-linked glycopeptides commonly identified in lung

cancer and control sera. The results were promising and allowed high reproducibility in peak area measurements during repeated MS-MS of targeted N-linked glycopeptides.

CONCLUSIONS

In this study, lectin−glycopeptide complexes were separated by size using AF4. This is an integrated strategy for simultaneously enriching targeted N-linked glycopeptides from proteolytic peptide mixtures by utilizing the specific affinity of lectins for glycans to form lectin−glycopeptide complexes. The present method provides simultaneous removal of nonglycosylated peptides during AF4 separation since typical peptides (<10 kDa) can readily be washed off through a channel membrane, which removes an additional depletion step. While the amount of lectins used for lectin−glycopeptide complexation in this work is similar to those of the conventional method, each single AF4 run requires only a few micrograms of lectins to retrieve about a half of a microgram of glycopeptides which are sufficient for the nLC-ESI-MS-MS run and the current size sorting method can surpass the weak points such as carry-over issues and low reproducibility. While a semiquantitative analysis of targeted N-linked glycoproteins from lung cancer samples was developed here, it is also necessary to implement an independent quantification method using an isotope-dilution MS (ID-MS) with absolute quantification (AQUA) of targeted peptide or enzyme-based ^{18}O-labeling. This will enhance the confidence level in quantitation to allow accurate determination of the relative regulation of glycoproteins. Unfortunately, absolute quantification analysis was not carried out in this study since the current work was focused on AF4 as a dual lectin-based size sorting and enrichment method. AF4 with multiple lectin affinity is suggested as an alternative method for the selective isolation of targeted N-linked glycopeptides and to quantitatively analyze the relative regulation of target biomarkers in relation to clinical stages.

ASSOCIATED CONTENT

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*D.K.: address, Center for Bioanalysis, Korea Research Institute of Standards and Science, 1 Doryong-Dong, Yuseong-Gun, Daejeon, 305-340, Korea; phone, (82) 42 868 5160; fax, (82) 42 868 5801; e-mail, djkang@kriss.re.kr. M.H.M.: address, Department of Chemistry, Yonsei University, Seoul, 120-749, Korea; phone, (82) 2123 5634; fax, (82) 2 364 7050; e-mail, mhmoon@yonsei.ac.kr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

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 (2011K000890) and in part by grant NRF-2011-0016438. S.-K.K. is appreciative of the support from the Basic Research Projects of KRISS, "Development of Protein Measurement Standards."

REFERENCES

(1) Dell, A.; Morris, H. R *Science* **2001**, *291*, 2351−2356.
(2) Marino, K.; Bones, J.; Kattla, J. J.; Rudd, P. M. *Nat. Chem. Biol.* **2010**, *6*, 713−723.
(3) Helenius, A.; Aebi, M. *Annu. Rev. Biochem.* **2004**, *73*, 1019−1049.
(4) Dube, D. H.; Bertozzi, C. R. *Nat. Rev. Drug Discovery* **2005**, *4*, 477−488.
(5) Hagglund, P; Bunkenborg, J; Elortza, F.; Jensen, O. N.; Roepstorff, P. *J. Proteome Res.* **2004**, *3*, 556−566.
(6) Bause, E.; Legler, G. *Biochem. J.* **1981**, *195*, 639−644.
(7) Morelle, W.; Michalski, J. C. *Nat. Protoc.* **2007**, *2* (7), 1585−1602.
(8) Wang, H.; Wong, C.-H.; Chin, A.; Taguchi, A.; Taylor, A.; Hanash, S.; Sekiya, S.; Takahashi, H.; Murase, M.; Kajijara, S.; Iwamoto, S.; Tanaka, K. *Nat. Protoc.* **2011**, *6* (3), 253−269.
(9) Lazar, L. M.; Lazar, A. C.; Cortes, D. F.; Kabulski, J. L. *J. Electrophor.* **2011**, *32*, 3−13.
(10) Yang, Z.; Harris, L. E.; Palmer-Toy, D. E.; Hancock, W. S. *Clin. Chem.* **2006**, *52*, 1897−1905.
(11) Madera, M.; Mechref, Y.; Novotny, M. V. *Anal. Chem.* **2005**, *77* (13), 4081−4090.
(12) Gonzalez-Begne, M.; Lu, B.; Liao, L.; Xu, T.; Bedi, G.; Melvin, J. E.; Yate, J. R., III *J. Proteome Res.* **2011**, *10*, 5031−5046.
(13) Sun, B.; Ranish, J. A.; Utleg, A. G.; White, J. T.; Yan, X.; Lin, B.; Hood, L. *Mol. Cell. Proteomics* **2007**, *6* (1), 141−149.
(14) Hogan, J. M.; Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *J. Proteome Res.* **2005**, *4*, 628−632.
(15) Wu, S. L.; Huhmer, A. F.; Hao, Z.; Karger, B. L. *J. Proteome Res.* **2007**, *6* (11), 4230−4244.
(16) Zubarev, R. A. *Mass Spectrom. Rev.* **2003**, *75*, 3256−3262.
(17) Alvarez-Manilla, G.; Atwood, J., III; Guo, Y.; Warren, N. L.; Orlando, R.; Pierce, M. *J. Proteome Res.* **2006**, *5*, 701−708.
(18) Jebanathirajah, J.; Steen, H.; Roepstorff, P. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (7), 777−784.
(19) Jung, K.; Cho, W.; Regnier, F. E. *J. Proteome Res.* **2009**, *8* (2), 643−650.
(20) Kaji, H.; Yamauchi, Y.; Takahashi, N.; Isobe, T. *Nat. Protoc.* **2007**, *1* (6), 3019−3027.
(21) Larsen, M. R.; Hojrup, P.; Roepstorff, P. *Mol. Cell. Proteomics* **2005**, *4*, 107−119.
(22) Kang, D.; Ji, E. S.; Moon, M. H.; Yoo, J. S. *J. Proteome Res.* **2010**, *9*, 2855−2862.
(23) Li, Y.; Tao, S.-C.; Bova, G. S.; Liu, A. Y.; Chan, D. W.; Zhu, H.; Zhang, H. *Anal. Chem.* **2011**, *83*, 8509−8516.
(24) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21*, 255−261.
(25) Giddings, J. C. *Science* **1993**, *260*, 1456−1465.
(26) Rambaldi, D. C.; Reschiglian, P.; Zattoni, A.; Johan, C. *Anal. Chim. Acta* **2009**, *654*, 64−70.
(27) Kang, D.; Moon, M. H. *Anal. Chem.* **2005**, *77*, 4207−4212.
(28) Kang, D.; Oh, S.; Ahn, S.-M.; Lee, B.-H; Moon, M. H. *J. Proteome Res.* **2008**, *7* (8), 3475−3480.
(29) Kim, K. H.; Moon, M. H. *Anal. Chem.* **2009**, *81* (4), 1715−1721.
(30) Kang, D.; Yoo, J. S.; Kim, M. O.; Moon, M. H. *J. Proteome Res.* **2009**, *8* (2), 982−991.
(31) Kang, D; Moon, M. H. *Anal. Chem.* **2006**, *78*, 5789−5798.
(32) Kang, D.; Oh, S.; Reschiglian, P.; Moon, M. H. *Analyst* **2008**, *133*, 505−515.
(33) Reschiglian, P.; Moon, M. H. *J. Proteomics* **2008**, *71* (3), 265−276.
(34) Kim, K. H.; Moon, M. H. *Anal. Chem.* **2011**, *83* (22), 8652−8658.
(35) Treuheit, M. J.; Costello, C. E.; Halsall, H. B. *Biochem. J.* **1992**, *283*, 105−112.
(36) Kang, D.; Nam, H.; Kim, Y. -S.; Moon, M. H. *J. Chromatogr., A* **2005**, *1070*, 193−200.
(37) Goldberg, D.; Bern, M.; Parry, S.; Sutton-Smith, M.; Panico, M.; Morris, H. R.; Dell, A. *J. Proteome Res* **2007**, *6*, 3995−4005.

(38) Shen, Z.; Want, E. J.; Chen, W.; Keating, W.; Nussbaumer, W.; Moore, R.; Gentle, T. M.; Siuzdak, G. *J. Proteome Res.* **2006**, *5*, 3154−3160.
(39) Gao, B.-B.; Chen, X.; Timothy, N.; Aiello, L. P.; Feener, E. D. *J. Proteome Res.* **2008**, *7* (6), 2516−2525.
(40) Martinez-Hernandez, R.; Montes, S.; Higuera-Calleja, J.; Yescas, P.; Boll, M. C.; Diaz-Ruiz, A.; Rios, C. *Neurochem. Res.* **2011**, *36* (11), 2111−2115.
(41) Lafuente, M. J.; Casterad, X.; Laso, N.; Mas, S.; Panades, R.; Calleja, A.; Hernandez, S.; Turuguet, D.; Ballesta, A.; Ascaso, C.; Lafuente, A. *Toxicol. Lett.* **2002**, *136*, 9−17.
(42) Someya, M.; Sakata, K. I.; Matsumoto, Y.; Kamdar, R. P.; Kai, M.; Toyota, M.; Hareyama, M. *Br. J. Cancer* **2011**, *104* (11), 1724−1729.
(43) Ueda, K.; Katagiri, T.; Shimada, T.; Irie, S.; Sato, T. A.; Nakamura, Y.; Daigo, Y. *J. Proteome Res* **2007**, *6*, 3475−3483.
(44) Fishelson, Z.; Donin, N.; Zell, S.; Schultz, S.; Kirschfink, M. *Mol. Immunol.* **2003**, *40*, 109−123.
(45) Doles, J.; Oliver, T. G.; Cameron, E. R.; Hsu, G.; Jacks, T.; Walker, G. C.; Hemann, M. T. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *107* (48), 20786−20791.