Efficient enrichment of glycopeptides using metal–organic frameworks by hydrophilic interaction chromatography†

Yongsheng Ji, a,b Zhichao Xiong, c Guang Huang, a Jing Liu, a Zhang Zhang, a Zheyi Liu, a Junjie Ou, a Mingliang Ye a and Hanfa Zou a*

Selective enrichment of glycopeptides from complicated biological samples is critical for glycoproteomics to obtain the structure and glycosylation information of glycoproteins using mass spectrometry (MS), which still remains a great challenge. Hydrophilic interaction chromatography (HLIC)-based strategies have been proposed for selective isolation of glycopeptides via the interactions between the glycan of glycopeptides and the matrices. However, the application of these methods is limited by the medium selectivity of HLIC matrices. In this study, hydrophilic metal–organic frameworks (MOFs) were fabricated and used as a HLIC matrix. The cross-linked CD-MOFs (LCD-MOFs) were facilely prepared with γ-cyclodextrin as ligand and possessed nano-sized cubic structure, superior hydrophilicity, and bio-compatibility. The LCD-MOFs performance for the selective enrichment of glycopeptides from the complex biological samples were investigated with a digested mixture of human immunoglobulin G (IgG) that was used as standard samples. In the selectivity assessment, the non-glycopeptides causing ion suppression to the glycopeptides were effectively removed, the signal of glycopeptides were enhanced significantly by LCD-MOFs, and twenty glycopeptides were identified with 67 fmol of IgG digest. In addition, the resulting LCD-MOFs demonstrated the lower detection limit (3.3 fmol) with a satisfactory recovery yield (84–103%) for glycopeptide enrichment from a digest of IgG. Furthermore, a promising protocol was developed for the selective enrichment of glycopeptides from mouse liver, and 344 unique N-glycosylation sites that mapped to 290 different glycoproteins were identified in a single MS run. The results clearly demonstrated that when used in a HLIC matrix, LCD-MOFs have great potential for identifying and enriching low-abundant glycopeptides in complex biological samples.

Introduction

Glycosylation, one of the most important post-translational modifications (PTMs) of proteins, produces abundant and diverse glycoproteins that participate in many key biological processes such as cell adhesion, receptor activation, and signal transduction.1 Aberrant glycosylation is a hallmark of cancer, and various biomarkers of diseases are glycoproteins.2 Thus, developing new methodologies to discern glycoproteins will not only elucidate the physiological mechanism of living organisms but also accelerate biomarker discovery for disease diagnosis or prognosis. Currently, the primary strategy is to implement enzymatic digestion of the glycoprotein followed by glycopeptide sequencing through mass spectrometry (MS) for the comprehensive characterization of glycoproteins.3 However, despite the advancements in MS technology, direct MS analysis of glycoproteins is still a challenge due to the inherent low abundance, the low signal response, and the severe ion suppression caused by the co-existence of non-glycosylated peptides. Therefore, an efficient enrichment platform is indispensable for the comprehensive analysis of glycoproteins by MS.

Several effective strategies including lectin affinity chromatography (LAC),4,5 hydrazide chemistry,6,7 boronic acid chemistry,8–10 and hydrophilic interaction chromatography (HLIC)11–16 have been developed for the enrichment of glycopeptides and glycopeptides. Among them, LAC based on the affinity interaction between lectin and specific glycan moieties is widely used, but the weak affinity and biased collection
restrict its application for the global analysis of glycoproteins. Hydrazide chemistry shows high selectivity through covalent bond formation. Unfortunately, the sample complexity is increased by the oxidation of carbohydrate cis-diol groups, and the glycan structure cannot be acquired due to the destruction and removal of the glycan segments. Boronic acid chemistry can capture glycoproteins or glycopeptides with intact glycan structures through reversible ester formation between boronic acid and cis-diol groups of glycan. However, there is insufficient selectivity and sensitivity for applications employing complex samples.

Recently, HILIC enrichment methods depending on the partition mechanisms between the organic solvent and the aqueous layer of the HILIC stationary phase have gained increasing popularity for the separation and purification of glycopeptides. HILIC-based tactics are advantageous because of their simple operation process, broad glycan specificity, and good compatibility with MS analysis. A variety of hydrophilic matrices such as Sepharose, cellulose, zwitterionic polymer, and click saccharide have been exploited for the selective extraction of glycopeptides. In previous works, carbohydrate-based sorbents displayed superior performances; however, the preparation protocol often suffers from harsh conditions, multistep reactions, and a tedious operation. Hence, cultivating a facile platform for producing carbohydrate-associated materials will involve a superior HILIC technique utilizing designed matrices for glycopeptide enrichment without difficulty.

Metal–organic frameworks (MOFs), a fascinating class of crystalline materials consisting of organic linkers with bridging organic ligands and metal ions, have been utilized in gas adsorption, catalysis, drug delivery, and separation with the advantages of high surface area, uniformly structured cavities, and available modification. Recently, MIL-101 and magnetic MOFs have been used as sorbents to enrich peptides from biological samples. However, it is still a challenge to utilize MOFs for the enrichment of peptides or proteins from biological samples. To the best of our knowledge, MOF-related enrichment of glycopeptides has not been reported. Generally, the vast majority of MOFs previously mentioned are composed of organic struts and transition metals, which limit their application for purifying glycoproteins and glycopeptides. Surprisingly, amazing MOFs with γ-cyclodextrin (γ-CD) as the ligand were easily manufactured under mild experimental conditions. Cyclodextrins, which are cyclic oligosaccharides from starch, are composed of six to eight glucose units bonded through α-(1,4) linkages. Owing to their truncated cone structure with hydrophilic rims and a hydrophobic cavity, CD-modified matrices are extensively involved in separation science and are widely used for the enrichment of glycopeptides via HILIC. A promising alternative for the selective enrichment of glycoproteins is to combine CD-MOFs with HILIC.

Herein, a MOF-based HILIC strategy was planned for the highly efficient and selective enrichment of glycopeptides from complicated biological samples. To illustrate the proof-of-concept, CD-MOFs were selected as adsorbents. MOFs utilizing γ-CD, an inexpensive and green feedstock as the ligand, were facely prepared. The as-prepared composite with a cubic structure possessed superior hydrophilicity, bio-compatibility, and ability to selectively enrich glycopeptides. The ability of CD-MOFs to selectively enrich glycopeptides from a standard digested mixture was tested, and then the desired platform was applied for analysis of N-linked glycopeptides in mouse liver. The results indicated that the developed approach possesses easy fabrication, high specificity, and expected capability in the selective enrichment of glycopeptides from complicated biological samples.

Materials and methods

Materials

Ethylene glycol diglycidyl ether (EGDE) and γ-cyclodextrin (γ-CD) were obtained from J&K Scientific Ltd. (Beijing, China). Methanol, acetonitrile (ACN), trifluoroacetic acid (TFA), and formic acid (FA) were supplied by Merck (Darmstadt, Germany). Human serum immunoglobulin G (human IgG), human α1-acid glycoprotein (AGP), human serotransferrin (TRF), bovine fetuin (BF), chick ovalbumin (COV), trypsin (TPCK treated), potassium hydroxide (KOH), dithiothreitol (DTT), iodoacetamide (IAA), 2,5-dihydroxybenzoic acid (DHB), cetyltrimethylammonium bromide (CTAB), concentrated ammonia aqueous solution (NH3·H2O, 28–30 wt%), and sodium bicarbonate (NaHCO3) were purchased from Sigma-Aldrich (St. Louis, MO, USA). PNGase F was from New England Biolabs (Ipswich, MA). Pure water (18.4 MΩ cm) used in all experiments was purified by a Milli-Q system (Millipore, Milford, MA, USA). All other chemicals were of analytical grade.

Preparation of CD-MOFs

CD-MOFs were synthesized according to a previous protocol with minor revision. γ-Cyclodextrin (163 mg, 0.126 mmol) was dissolved in 5.0 mL of 200 mM KOH aqueous solution in a glass bottle. The bottle was put into a jar, and methanol was allowed to vapor-diffuse into this solution at room temperature for 36 h. Then, the solution was transferred to a tube, and 5.0 mL of methanol containing CTAB (40 mg, 0.110 mmol) was added. The homogenous solution was incubated at room temperature for 3 h. The solution was then centrifuged at 1000 rpm for 3 min and the supernatant was decanted. The obtained crystals were washed with ethanol to remove excess reagents, and then dried under vacuum at room temperature for 12 h.

The dried CD-MOFs (40 mg) were added to 5 mL of anhydrous ethanol containing EGDE (7.5 mmol), and the mixture was kept at 65 °C under nitrogen protection for 3 days. Then, the mixture was centrifuged at 1000 rpm for 3 min. The white products were collected and washed with ethanol and water. The cross-linked CD-MOFs (LCD-MOFs) were dried under vacuum at room temperature for 12 h.

Material characterization

Transmission electron microscopy (TEM) images were obtained with a JEOL JEM-2000 EX transmission electron microscope
(JEOL, Tokyo, Japan). Fourier transform infrared spectroscopy (FTIR) characterization was performed on a Thermo Nicolet 380 spectrometer using KBr pellets (Nicolet, Wisconsin, USA). The water contact angles were measured with an OCA20 contact angle system (Dataphysics, Germany) at ambient temperature.

Tryptic digests of proteins

For tryptic digestion, 1 mg of human IgG was dissolved in 1 mL of NH4HCO3 solution (50 mM, pH = 8.3) and denatured by boiling for 15 min. After that, the samples were reduced with 20 mM DTT at 60 °C for 1 h and alkylated by 7.2 mg IAA at room temperature in the dark for 40 min. The solution was incubated with trypsin at an enzyme–protein ratio of 1 : 25 (w/w) at 37 °C for 16 h. The tryptic digests were stored at −20 °C until further use.

Proteins from mouse liver were extracted following a procedure that was previously described.46 Briefly, the liver was cut into pieces and suspended in ice-cold homogenization buffer consisting of 8 M urea, 4% CHAPS (w/v), 65 mM DTT, 1 mM EDTA, 0.5 mM EGTA, a mixture of protease inhibitor (1 mM PMSF, 0.2 mM NaN3, 1 mM NaF) and 40 mM Tris–HCl at pH 7.4. Then, the samples were homogenized by a Potter-Elvejhem homogenizer, sonicated at 100 W × 30 s and centrifuged at 25 000g for 1 h in turn. The total liver proteins in the supernatant were recovered by precipitation and lyophilization. The obtained proteins were dissolved in denaturing buffer (8 M urea, 50 mM Tris–HCl at pH 8.3). The protein concentration was measured by Bradford assay. Thereafter, the proteins were digested with trypsin. The protein mixture (1 mg) was reduced by DTT (1 M, 20 μL) at 60 °C for 1 h and alkylated by IAA (7.2 mg) in the dark at room temperature for 45 min. The solution was diluted to 1 M urea with Tris–HCl solution (50 mM, pH = 8.3) and digested by trypsin at 37 °C for 16 h at an enzyme–protein ratio of 1 : 25 (w/w). The tryptic digests were desalted and lyophilized for the next use. A five-glycoprotein mixture (IgG, AGP, TRF, BF, COV) was created and digested.

Enrichment of glycopeptides by LCD-MOFs

An amount of 150 μg LCD-MOFs were washed with loading buffer (ACN/H2O/TFA, 88 : 11.9 : 0.1, v/v/v) and dispersed in 200 μL of loading buffer containing 400 ng of IgG tryptic digests. After incubation for 30 min at room temperature, the suspension was centrifuged, and the supernatant was discarded by Eppendorf pipette, followed by rinsed three times with loading buffer (100 μL) to remove the non-glycopeptides. Then the captured glycopeptides were eluted by eluting buffer (20 μL, ACN/H2O/TFA, 30 : 69.9 : 0.1, v/v/v) under gentle shaking at room temperature for 10 min. The collected peptides were analyzed by MALDI-TOF MS.

For the glycopeptide enrichment from tryptic digests of mouse liver, 100 μg of the digests was dissolved in 200 μL loading buffer, incubated with 1.0 mg of LCD-MOFs for 30 min, and subsequently washed three times with 200 μL of loading buffer. Then, the trapped glycopeptides were eluted three times with 50 μL of eluting buffer under mild vortexing for 10 min, and the elution was collected and lyophilized. The obtained glycopeptides were redissolved in 10 mM NH4HCO3, and 100 units of PNGase F was added. After the mixture was incubated at 37 °C for overnight to remove the glycan moieties. The reaction was terminated by heating to 100 °C for 10 min, and then the mixture was lyophilized and redissolved prior to analysis by nano LC-MS/MS. The enrichment of glycopeptides from five-glycoproteins mixture was carried out as mentioned above.

Recovery estimation of glycopeptide enrichment

The recovery yield of glycopeptide enrichment can be achieved by stable isotope dimethyl labeling according to a previous report.47 Two aliquots of the same amounts of human IgG digests were firstly labeled with light and heavy isotopes. The heavy-tagged digests were enriched with LCD-MOFs according to the above-mentioned procedure, and the resulting elution was spiked into the light-tagged digests. The combined mixture was re-enriched with LCD-MOFs, and the eluted fraction was directly analyzed by MALDI-TOF MS. The recovery was calculated by the peak intensity ratio of heavy isotope-labeled glycopeptides to the light isotope-labeled glycopeptides.

Mass spectrometry analysis

All MALDI-TOF mass analyses were performed on an AB Sciex 5800 MALDI-TOF/TOF mass spectrometer (AB Sciex, CA) equipped with a pulsed Nd/YAG laser at 355 nm. The sample solution (0.5 μL) was deposited on the MALDI plate and dried at room temperature, and then 0.5 μL of DBH (25 mg mL−1, ACN/ H2O/H2PO4, 70 : 29 : 1, v/v/v) matrix solution was loaded for MS performance.

A nano-LC-MS/MS system consisting of an Accela 600 HPLC (Thermo, San Jose, CA), LTQ-Orbitrap Velos mass spectrometer (Thermo, San Jose, CA) and a homemade spray tip packed with C18 AQ beads (3 μm, 120 Å, Daison, Osaka, Japan) was equipped for the analysis of peptides. Buffer A (H2O/FA, 99.9 : 0.1, v/v) and buffer B (ACN/FA, 99.9 : 0.1, v/v) were provided for the mobile phase, and the gradient elution was programmed from 5% to 35% of buffer B in 120 min with a flow rate of 150 nL min−1. The MS/MS spectra were operated in a data-dependent collision induced dissociation (CID) mode, the full MS was acquired from m/z 400 to 2000 with a resolution of 60 000, the collision energy was 35.0%, and the activation time was 10 ms. The 20 most intense ions were selected for MS/MS. A lyophilized sample of 100 μg was redissolved in 50 μL of H2O/FA (99.9 : 0.1, v/v) for the nano LC-MS/MS analysis.

Database search and data analysis

All the LC-MS/MS raw data files were searched with MASCOT software (version 2.3) against the target database of five glycoprotein or the IPI mouse database (v3.80). For peptide identification, a mass tolerance of 20 ppm was permitted for parent ions and 0.5 Da for fragment ions, with allowance for two missed cleavages in the trypsin digests. Cysteine carboxyamidomethylation was set as a static modification of 57.0215 Da. The variable modifications of methionine oxidation and asparagine deamination are 15.9949 and 0.9840 Da, respectively.
Results and discussion

Preparation and characterization of CD-MOFs

The desired CD-MOFs were facilely prepared by incubation of γ-CD and 8.0 equiv. of KOH in water under methanol vapor. Cubic crystals were formed between maltosyls of CD and K ions. After crystal growth for 36 h, the microcrystals were removed, and methanol containing CTAB was added to the mother liquor for accelerating the second crystal growth. Then, the cross-linking reaction between the hydroxyl groups of each γ-CD in the CD-MOFs was accomplished by EGDE possessing two epoxy groups. The morphologies of the obtained LCD-MOFs were characterized by TEM, and the photographs are shown in Fig. 1 (a) and (b). The cubic particle size is approximately 200–300 nm, and the LCD-MOFs display evident cross-linking compared to the CD-MOFs. The hydrophilicity of the prepared materials was tested by the water contact angles. As shown in Fig. 1 (c) and (d), the angle of the LCD-MOFs distinctly increases compared to CD-MOFs, indicating that the cross-link process was successful and that there is sufficient hydrophilic ability of LCD-MOFs for their successful application in HILIC. The products were also determined by FT-IR and the spectra are listed in Fig. 1 (e). The typical peaks of 2925 and 1125 cm\(^{-1}\) are ascribed to the C–H and C–O–C stretching vibration, respectively.\(^{25}\) The strong adsorption band within 3200–3600 cm\(^{-1}\) is attributed to the hydroxyl groups of γ-CD. There are no obvious differences between the spectrum of CD-MOFs and LCD-MOFs owing to the similar composition. The characterizations reveal that the designed products possess uniformly nano-sized structure and superior hydrophilicity.

Glycopeptide enrichment from tryptic digest of human IgG

To evaluate the enrichment capacity of LCD-MOFs, a tryptic human IgG digest was employed as the test standard sample. The protocol for glycopeptide enrichment is illustrated in Scheme 1. The tryptic digest was incubated with LCD-MOFs for capturing the glycopeptides, followed in turn by the sorbents, which were collected and rinsed to remove non-glycopeptides, and then the captured glycopeptides on the LCD-MOFs were eluted for MALDI-TOF MS or nano LC-MS/MS analysis. Generally, it is assumed that the separation by HILC primarily relies on the partition mechanisms between the organic solvent and the aqueous layer on a hydrophilic surface of a solid matrix.\(^{3,29,30}\) Typically, a hydration layer around polar matrices is formed when a loading buffer with a relatively low water content is used. More hydrophilic glycopeptides are concentrated in the water layer, and more hydrophobic non-glycopeptides remain in the organic solvent, which contributes to the multivalent intermolecular interactions such as hydrogen bonding and electrostatic and dipolar forces.\(^{3,29,30}\) The water content of the loading buffer is the key factor for the selective enrichment of glycopeptides. Furthermore, the selectivity can be improved by adding an ion-pairing reagent to the loading buffer, which increases the hydrophobicity of non-glycopeptides.\(^{3,12}\) Therefore, the ratio of ACN : water and TFA percentage in the loading buffer were investigated. As shown in Fig. 2, the small variation of the ACN : water ratio greatly influenced the enrichment of glycopeptides, but an increase in the TFA volume fraction from 0.1 to 3.0 could not obviously intensify the enrichment efficiency. The incubation time and the amount of sorbents were optimized to gain a higher enrichment efficiency, and the results are included in Fig. 2. According to the experimental results, a loading buffer of 88% ACN containing 0.1% TFA (v/v), 150 µg of LCD-MOFs, and incubation for 30 min were adopted for the next experiment.

The selectivity of LCD-MOFs for glycopeptide enrichment was also tested. Fig. 3 (a) and (b) shows MALDI mass spectra for the direct analysis of 67 fmol human IgG digest and treatments by LCD-MOFs. Only two glycopeptides were barely detectable, and non-glycopeptide peaks dominated the MS spectrum. However, after enrichment with LCD-MOFs, the non-glycopeptides were nearly removed, and the glycopeptide signals were significantly enhanced, indicating that the unique structures and superior hydrophilicity of LCD-MOFs contributed to the performance of glycopeptide enrichment via multivalent
hydrophilic interactions. Furthermore, the eluted fraction after enrichment was deglycosylated by PNGase F, and the obtained product was further analyzed by MALDI-TOF MS (Fig. 3(c)). It can be seen that the MS peaks of presumed glycopeptides disappeared, which confirmed that the peaks in Fig. 3(b) can be attributed to glycopeptides. In addition, twenty glycopeptides from the tryptic digest of human IgG were identified based on the MS data, and the detailed sequences are listed in Table S1.†

Detection sensitivity of glycopeptides from a tryptic human IgG digest

The detection sensitivity of glycopeptides was measured for assessing the enrichment efficiency of LCD-MOFs. As shown in Fig. 4, low amounts of a tryptic human IgG digest (13.4 pmol, 6.7 fmol and 3.3 fmol) were tested by MALDI-TOF MS. Four glycopeptides with $S/N > 3$ could still be detected while the amount of tryptic human IgG digest was as low as 3.3 fmol (Fig. 4(c)), which is lower than that enriched by zwitterionic HILIC materials (10 fmol) and hydrophilic silica-based click maltose (30 fmol). Such satisfactory results demonstrated that LCD-MOFs can be used for glycopeptide enrichment with high sensitivity.

Recovery yield test of LC-MOFs for glycopeptide enrichment

Stable-isotope dimethyl labeling was used to study the recovery yield of LCD-MOFs for glycopeptide enrichment. One of two equivalent digests labeled with light and heavy isotopes was enriched with LCD-MOFs, and the resulting eluted fraction was combined with the other, followed in turn by enrichment again and then analysis by MALDI-TOF MS. The recovery was calculated by the peak intensity ratio of the heavy-tagged glycopeptides to light-tagged glycopeptides. Fig. 5 presents the recovery yields of glycopeptides ranging from 84% to 103%, which confirmed that LCD-MOFs possess great potential for the enrichment of glycopeptides.

Glycopeptide enrichment from a five-glycoprotein mixture and mouse liver sample

LCD-MOFs were employed to enrich glycopeptides from a complex sample containing five standard glycoproteins. After enrichment and deglycosylation, the resulting samples were analyzed by nano LC-MS/MS, and the results are listed in Table S2.† It can be seen that 17 out of 23 unique N-glycosylation sites annotated in the UniProt database were identified from 3 μg of the five-glycoprotein mixture. Five target glycoproteins were ascertained, and most of unique sites were covered. The above results revealed that LCD-MOFs have a remarkable ability for glycopeptide enrichment.

With the superior efficiency of enrichment glycopeptides in mind, LCD-MOFs were finally applied for the analysis of N-glycoproteome in a complex mouse liver sample. A tryptic mouse liver digest of 100 μg was incubated with LCD-MOFs for enriching glycopeptides. The eluted glycopeptides were then
and LCD-MOFs. LCD-MOFs were further applied to analyze the help of the multivalent interactions between glycopeptides for enrichment of glycopeptides from a glycoprotein digest with selectivity, high detection sensitivity, and satisfactory recovery and sensitive HILIC MOFs for glycoproteome analysis. This work is demonstrated that LCD-MOFs possessed great potential for the enrichment of low-abundant glycopeptides and characterization of glycoproteome in real biological samples. These results proved that LCD-MOFs show a superior capability for the characterization of glycoproteome in complex biological samples.

Conclusions

In summary, uniformly cubic LCD-MOFs were facilely prepared. The resulting nano-sized LCD-MOFs composites possessed the desired hydrophilicity due to the hydrophilic γ-CD, which acted as a ligand. It was confirmed that LCD-MOFs showed superior selectivity, high detection sensitivity, and satisfactory recovery for enrichment of glycopeptides from a glycoprotein digest with the help of the multivalent interactions between glycopeptides and LCD-MOFs. LCD-MOFs were further applied to analyze N-glycoproteome in a complex mouse liver sample, and the results demonstrated that LCD-MOFs possessed great potential for the enrichment of low-abundant glycopeptides and characterization of glycoproteome in real biological samples. This work is expected to open up a new promising strategy to design efficient and sensitive HILIC MOFs for glycoproteome analysis.

Acknowledgements

This work is supported by the China State Key Basic Research Program Grant (2013CB911202, 2012CB910604), the Creative Research Group Project of NSFC (21321064), the National Natural Science Foundation of China (21275142, 21235006, 81161120540), the National Key Special Program on Infectious Diseases (2012ZX10002009-011), and the Analytical Method Innovation Program of MOST (2012IM030900) to H. F. Zou. We also gratefully acknowledge the support of the Key Project of Henan Province Education Department Science and Technology Research (12A350002) and the Doctoral Research Fund of Henan Chinese Medicine (BSJJ-2010-23).

Notes and references