# Label-Free Discovery Array Platform for the Characterization of Glycan Binding Proteins and Glycoproteins

Christopher John Gray, Antonio Sánchez-Ruiz, Ivana Šardzíková, Yassir A. Ahmed, Rebecca Louise Miller, Juana E. Reyes Martinez, Edward Pallister, Kun Huang, Peter Both, Mirja Hartmann, Hannah N. Roberts, Robert Šardzik, Santanu Mandal, Jeremy E. Turnbull, Claire E. Eyers, and Sabine L. Flitsch

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

# Label-Free Discovery Array Platform for the Characterization of Glycan Binding Proteins and Glycoproteins

Christopher J. Gray,[†][‖] Antonio Sánchez-Ruíz,[†][‖] Ivana Šardzíková,[†] Yassir A. Ahmed,[‡] Rebecca L. Miller,[‡] Juana E. Reyes Martinez,[#] Edward Pallister,[†] Kun Huang,[†] Peter Both,[†] Mirja Hartmann,[†§] Hannah N. Roberts,[†] Robert Šardzík,[†][φ] Santanu Mandal,[†] Jerry E. Turnbull,*[‡] Claire E. Eyers*[‡] and Sabine L. Flitsch*[†]

[†] School of Chemistry and Manchester Institute of Biotechnology. The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK.

[‡] Department of Biochemistry, Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool, L69 7ZB, UK.

[#] Departamento de Biología, División de Ciencias Naturales y Exactas, Universidad de Guanajuato, Col. Noria Alta S/N, Guanajuato 36050, México.

**ABSTRACT:** The identification of carbohydrate-protein interactions is central to our understanding of the roles of cell-surface carbohydrates (the glycocalyx), fundamental for cell-recognition events. Therefore there is a need for fast high-throughput biochemical tools to capture the complexity of these biological interactions. Here we describe a rapid method for qualitative label-free detection of carbohydrate-protein interactions on arrays of simple synthetic glycans, more complex natural glycosaminoglycans (GAG) and lectins/carbohydrate binding proteins using MALDI-ToF mass spectrometry. The platform can unequivocally identify proteins that are captured from either purified or complex sample mixtures, including biofluids. Identification of proteins bound to the functionalized array is achieved by analyzing either the intact protein mass, or, after on-chip proteolytic digestion, the peptide mass fingerprint and/or tandem mass spectrometry of selected peptides, which can yield highly diagnostic sequence information. The platform described here should be a valuable addition to the limited analytical toolbox that is currently available for glycomics.

A diverse variety of complex oligosaccharides coat the surface of all mammalian cells via membrane embedded proteins or lipids. A key function of these complex oligosaccharide structures is to regulate intracellular recognition events including cell adhesion through specific adherence of glycan-binding proteins (GBPs). These complex oligosaccharides are also targeted during pathogenic infection.[1-4] Aberrant glycosylation of native species has been observed in many diseases, notably a variety of cancers, acute pancreatitis and muscular dystrophies.[5-8] A robust, high-throughput analytical technique to characterize carbohydrate-protein binding partners and ultimately for use in screening for biomarkers of disease is therefore highly sought-after. Most current high-throughput strategies require the incorporation of chemical label, which can be challenging for complex mixtures such as clinical samples.[9-11] Label-free detection techniques are also desirable, especially those that allow unequivocal identification of unknown sample components. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) mass spectrometry (MS) is an analytical technique perfectly suited for this application, as it is label-free, rapid, sensitive and capable of providing structural information. The application of MALDI-ToF to the

study of glycoenzyme activity on arrays has been extensively studied by us[12-16] and others[17-19]. The identification of glycan-GBP partners using MALDI-ToF MS however has been little explored.[17, 20-22] The Reichardt group reported enrichment of urine proteins on an array of lectins, however are unable to characterize these bound species.[22] Unambiguous assignment of GBPs directly from complex mixtures of biofluids, would greatly facilitate glycobiology studies, although remains to be achieved. Here, we report investigation of glycan, glycosaminoglycan (GAG) and lectin arrays for the label-free identification of GBP-glycan interactions even from complex crude protein extracts. Using a single generic platform, MALDI-ToF MS(/MS) was used to detect and characterize purified GBPs, GAG binding proteins and glycoproteins, without the requirement of a chemical label (Figure 1). The incorporation of affinity-capture into the platform enables enrichment of low-level binding proteins from complex biofluids. Previously reported on-chip proteolysis[22] allowed characterization of these enriched proteins using a standard bottom-up proteomic strategy i.e. database searching of peak lists generated from both mass and tandem mass spectra. The utility and versatility of our platform has been used to investigate carbohydrate-protein interac-

tions in human milk samples, a biofluid whose glycan composition appears to be critical for the innate immune response of neonates[23, 24].

The present array platform consists of a chemically cleaned gold surface, on which a self-assembled monolayer (SAM) of long chain alkanethiols is formed, exploiting the relatively strong Au-S bond. These alkanethiols terminate in either $(PEG)_3OH$ ('spacer') or $(PEG)_6COOH$ ('linker') moieties. Terminal COOH groups are readily covalently functionalized with amino-functionalized glycans or lectins by standard EDC and PFP/NHS activation, whilst terminal PEG groups suppress non-specific protein adsorption.[25, 26] COOH 'linkers' could also be converted to the equivalent carbohydrazide allowing immobilization of underivatized oligosaccharides.[27] The density of carbohydrates or lectins presented on the array can be easily controlled by altering the ratio of linker and spacer. SAMs are eminently suited to studying bio-interactions as to some extent they mimic the fluidic mosaic assembly of cell surfaces.[25] Also, ethylene glycol linkers have recently been shown to be highly flexible, which allows for favorable presentation of terminal glycan ligands.[28] Conveniently, self-assembly of the alkanethiols and subsequent glycan derivatization can be quickly monitored by MALDI-ToF MS (acquisition ~2 s per spot) (Figure S1, S2 and S3).



**Figure. 1** Discovery tool-kit for carbohydrate-binding proteins: Self-assembled monolayers (SAMs) on a gold array are functionalized with glycans (left spot) or glycan-binding proteins (right spot) (a). When complex protein mixtures from biological samples are incubated on the array, glycan-binding proteins (left) or glycoproteins (right), respectively

can be enriched, detected and identified on the array by MALDI ToF MS(/MS) with or without on-target trypsin digestion. The 'linkers' and 'spacers' we use to immobilize these substrates are also depicted (b). EDC, PFP, NHS and GAG correspond to N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride, pentafluorophenol, N-hydroxysuccinimide and glycosaminoglycan respectively.

## Experimental Section

Commercially available starting materials and reagents were used without further purification. PEGylated alkane thiols HS-(CH2)17-(EG)3-OH (spacer) and HS-(CH2)17-(EG)6-CH2COOH (linker) were purchased from Prochimia Surfaces (Poland). N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC), N-hydroxysuccinimide (NHS), pentafluorophenyl (PFP), Betaine, trifluoroacetic acid (TFA) 2', 4', 6'-trihydroxyacetophenone monohydrate (THAP), sinapic acid (SA), α-cyano-4-hydroxycinnamic acid (CHCA) (purity > 99.5%) and trypsin (porcine pancreas proteomics grade) were purchased from Sigma-Aldrich. Lectins were purchased from Vector Labs (Burlingame, USA), dermatan sulfate, chondroitin sulfate A, 50 wt.% mw ~1300 polyethylenimine (PEI) heparin-binding EGF-like growth factor (HBEGF), tumor necrosis factor-alpha (TNF-α, fibroblast growth factor 2 (FGF2) and ovalbumin from chicken egg and bovine serum albumin fraction V (BSA) were purchased from Sigma-Aldrich. Most aminoethyl glycosides were prepared according to published procedures.[29] Heparin was purchased from Merk Millipore (USA) whereas dp2, dp4, dp8 dp14, dp18 and dp26 oligosaccharides were prepared from porcine mucosal heparin (Celsus, cat. no. PH-3005) by nitrous acid digestion followed by size exclusion chromatography.[30] Milk samples were acquired from a healthy volunteer mother. Aliquoted samples were stored at -40°C until analysis. Proteins were extracted from doubly defatted milk[31] using centrifugal filters (10 K, Millipore) in accordance to the manufacturer's instructions. All solvents employed were HPLC grade. Heat maps were generated in R using gplots.

## MALDI-ToF MS(/MS) and database searching

Gold AB plates (AB Sciex Ltd, USA) were loaded into the instrument using the MTB AB adapter (Bruker). MALDI-ToF mass spectra (1500 shots/spectrum) were recorded on a Bruker Ultraflex II instrument with a Smartbeam I laser in positive reflector ion mode (SAMs, glycans and peptides) or in positive linear mode (proteins). Tandem MS (4500 shots/spectrum) were recorded in positive LIFT ion mode with precursor ion selection of peptide ions prior to collisional dissociation. The laser energy was set at a value that did not result in loss of mass resolution (typically 25-40%). Data were analyzed and normalized using FlexAnalysis version 3.0 (Bruker). Positive reflector ion mode was calibrated between 700-3500 $m/z$ using a solution of peptide calibration mix II (Bruker Daltonics, Bremen) and
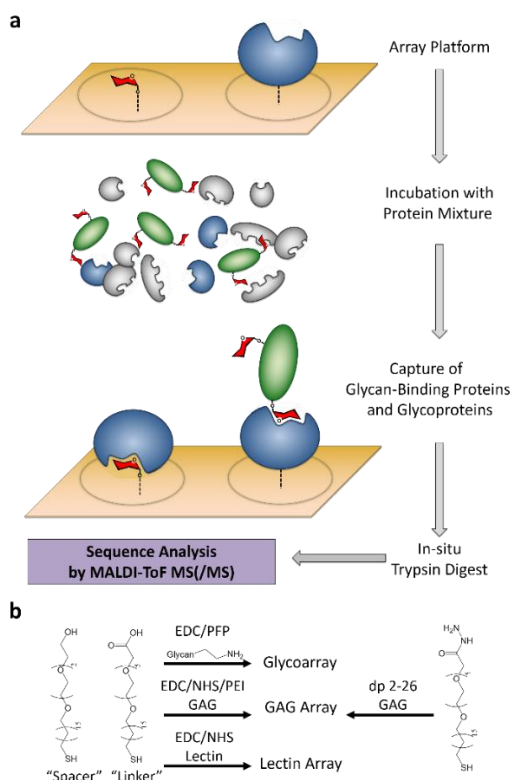
2

positive linear ion mode between 12000-70000 $m/z$ using a solution of ProteoMassTM Protein MALDI-MS Calibration Kit (Sigma-Aldrich) before each experiment. Tandem MS (LIFT) was calibrated to fragments from the peptide $P_{14}R$. Intact captured proteins and tryptic peptides were analyzed using 1 µL SA (20 mg/mL, in 7:3 (v/v) acetonitrile:water 0.1% TFA) and 0.5 µL CHCA (20 mg/mL sat., in 1:1 (v/v) acetonitrile:water 0.1% TFA) as matrices respectively. MS data acquisition was automated where possible using WARP-LC (Bruker) software (acquisition set to random walk, 12 shots per spot). Mascot was used to search the peak lists generated from peptide mass fingerprints and tandem mass spectra against the Swiss-Prot database (update 2013_12 until 2014_8) using the following parameters: singly charged tryptic peptides, 1 missed cleavage, variable oxidation of methionine, precursor ion tolerance of 100 ppm and product ion tolerance of 0.6 Da, unless otherwise stated. The taxonomy was defined as appropriate.

Heparin saccharide experiments were performed on a MALDI AXIMA-ToF (Shimadzu Biotech) discussed in detail within the electronic supplementary information section 3.2.

## Results and discussion

### Glycoarrays to study lectin specificity

Glycoarrays have been used extensively to study the interactions of fluorescently labeled lectins with immobilized glycan structures.[32, 33] Therefore, as a proof-of-concept we tested the platform with well-studied unlabelled lectins to establish that the binding results were comparable to those reported for standard glycoarrays that employ labeled lectins. The SAMs were functionalized with an array of natural aminoethyl derivatised saccharides including *N*-acetylglucosamine (Glc*NAc*), *N*-acetylgalactosamine (Gal*NAc*), mannose (Man), galactose (Gal), glucose (Glc), fucose (Fuc), xylose (Xyl), lactose (Lac) and N-acetyllactosamine (Lac*NAc*).[13, 29] The synthesis of the β-aminoethyl GalNAc glycosides is reported in the SI section 2.4 (Figure S4). Other glycosides were synthesized according to our previously published methods[29] other than α-aminoethyl Gal*NAc*, which was prepared according to the method of Ni et al.[34] α-2,3 and α-2,6 sialylated lactose were prepared enzymatically from lactose on-chip using trans-sialidase (TcTS) and a sialyl transferase (Pd2,6ST) respectively, circumventing challenging chemical syntheses.[14] These relatively simple glycans were chosen as they encompass many of the terminal mammalian glycans that have been characterized within natural samples. This includes the Tn antigen (α-Gal*NAc*) and the trisaccharide α-2,3 sialylated lactose representing the terminus of sialyl Le$^x$, both of which are known cancer biomarkers.[35] These arrays were probed with a range of commonly used purified plant lectins: ConA, WGA, LTL, SNA, VVA, GSL-I B$_4$, MAL I and MAL II, whose glycan specificities have been extensively studied.[36-40] The arrays were subsequently washed with aqueous buffer and specific binding of the intact lectins was determined using MALDI-ToF MS (Table 1, Figure 2a and S5).

Table 1 shows that there was minimal non-specific binding of lectins to the platform. In fact, the results presented corresponded broadly to affinities reported in the literature for each lectin: WGA was detected on the spots functionalized with Glc*NAc*, Gal*NAc*, Neu5Ac and interestingly Lac*NAc*; VVA bound to Gal*NAc* arrays and ConA bound to α-Man-presenting SAMs (Figure 2a). Also as expected, SNA was observed to specifically bind to α-2,6Neu5Ac terminated glycans rather than the similar α-2,3Neu5Ac motif. The α-Gal specific lectin GSL-I B$_4$ was observed to bind to β-Gal and Lac (Galβ-1,4Glc), which might be indicative of cross-reactivity with its closely related natural ligand. These results suggest that the lectins retained their affinity for the covalently immobilized glycans. We did not observe LTL binding to a single Fuc residue and has previously been reported to have a relatively low binding affinity towards a single fucose moiety and requires more of the global glycan composition which this particular substrate surface does not provide.[40] It is also surprising that neither MAL I or MAL II displayed binding towards their preferred residues Galβ-1,4-Glc*NAc* and Neu5Acα-2,3-Galβ-1,4-Glc*NAc* respectively. These lectins therefore either require more of the core glycan structure or its binding is inhibited by the presence of the SAM.[38]

The label-free nature of the present platform allows for multiplexed experiments, such as the study of lectin mixtures, without the need for differential labeling. As an extension of the initial studies, ratiometric mixtures of WGA and ConA were incubated with different glycan surfaces as above, resulting in lectin binding only to their literature reported glycan ligand (Figure S6). These results indicated that the platform is compatible with studying glycan-protein interactions from crude cell extracts or biofluids.

### Glycoarrays for identification of GBPs from mixtures

Moving from individual purified proteins or simple mixtures of known content to natural protein extracts posed a problem with our current methodology – it is impossible to unambiguously identify a protein by virtue of intact mass alone given the sheer complexity of protein content within a biological sample and the heterogeneity of post-translational modifications. Unambiguous identification of bound proteins was therefore facilitated by on-chip proteolysis to generate peptide fragments with known features (termini) that could be used both as diagnostic fingerprints (peptide mass fingerprinting)[41] and/or subjected to tandem MS (MS/MS) to elucidate the primary peptide sequence (Figure 1). Captured lectins were subjected to on-chip tryptic digestion and the resulting peptides were analyzed by MALDI-ToF MS(/MS).

3

**Table 1 Qualitative binding specificities of 11 immobilized glycans with 8 plant lectins using the array platform. Positive hits represent signals associated with that lectin being observed within the intact protein mass spectrum (Figure S5 shows selected spectra) unless otherwise stated. Observed binding is represented by √; no binding: x;. Signals were observed for ConA at ~26k Da and ~13k Da, WGA at 17 kDa, LTL at 58 kDa and 29 kDa, VVA at 29 kDa, GSL I-B$_4$ at 31 kDa, MAL I at 65 kDa and 32 kDa and MAL II at 30 kDa. SNA yielded poorer intact mass protein spectra (weak signal at 17 kDa and therefore were characterized after trypsin digestion yielding diagnostic peptides at $m/z$ 2421.2 and 2604.3. Green indicates agreement with current literature-derived binding specificities; red indicates disagreement with current knowledge. The blank is a self-assembled monolayer composed entirely of 'spacers' with no 'linker' thiols.**

| Lectin | Specificity | β-Glc | α-Man | β-Gal | β-GlcNAc | α-GalNAc | β-GalNAc | α-Fuc | β-Xyl | β-Lac | β-LacNAc | Neu5Acα2-3LacNAc | Neu5Acα2-3Lac | Neu5Acα2-6Lac | Blank |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Con A | α-Man/ α-Glc | x | √ | x | x | x | x | x | x | x | x | x | x | x | x |
| WGA | GlcNAc/Neu5Ac | √ | x | x | √ | √ | √ | x | x | x | √ | √ | √ | √ | x |
| LTL | α-Fuc | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| SNA | Neu5Acα2-6Gal(NAc) | x | x | x | x | x | x | x | x | x | x | x | x | √ | x |
| VVA | GalNAc | x | x | x | x | √ | √ | x | x | x | x | x | x | x | x |
| GSL-I B$_4$ | α-Gal | x | x | √ | x | x | x | x | x | √ | x | x | x | x | x |
| MAL I | Galβ1-4GlcNAc | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| MAL II | Neu5Acα2-3Galβ1-4GlcNAc | x | x | x | x | x | x | x | x | x | x | x | x | x | x |

Peaks corresponding to products arising due to autolysis of trypsin ($m/z$ 842 and 1045) were observed in all the samples analyzed; these ions served as useful reference points allowing internal mass calibration and relative quantification. Characterization of peptide precursor and product ions (generated by collision-induced dissociation of selected precursors), was aided by database (Swiss-Prot) searching using the Mascot search engine.[42] The information obtained from the intact mass spectrum, peptide mass fingerprint and tandem mass spectra combined enabled unambiguous characterization of the bound lectin(s). Figure 2 demonstrates this approach for ConA, which was observed to bind only to Man-functionalized arrays (Figure 2a). Trypsin proteolysis of the captured ConA yielded peptides covering 48% of the protein sequence (Figure 2b). Searching the $m/z$ peak list (mass fingerprint) against the Swiss-Prot database resulted in a significant hit for ConA. Furthermore, collision-induced dissociation of the precursor ion isolated at $m/z$ 2103 (Figure 2c) identified the peptide as DLILQGDATTGTDGNLELTR, a sequence unique to ConA, further confirming the identity of this protein. Having optimized the binding of these glycoarrays with purified standards, we then proceeded to analyze a complex whole wheat germ protein extract (see electronic supplementary information section 2.7 for preparation). MALDI-ToF MS of the crude protein mixture shows the protein heterogeneity within the extract (Figure S7a). After incubation with a GlcNAc presenting glycoarray, singly and doubly charged ions ($[M+H]^+$ = $m/z$ 17062; $[M+2H]^{2+}$ = $m/z$ 8526) corresponding to peaks expected for WGA are observed, consistent with the lectin's known specificity and our initial screen (Figure S7b). These results exemplify the potential application of our glycoarray to enrich and characterize GBPs even from complex biological mixtures.

## Glycosaminoglycan Arrays

Having demonstrated the successful application of this platform on relatively simple glycans, we tailored the surface to present more complex and physiologically relevant glycosaminoglycans (GAGs), such as heparin/heparin sulfate (HS), dermatan sulfate (DS) and chondroitin sulfate (CS). GAGs are a category of glycans composed entirely of hexosamine and hexose or hexuronic acid disaccharide repeating units. These repeating units can be sulfated, making them highly anionic, and therefore often difficult to purify and analyze. HS is present on the cell surface, in the extracellular matrix, and is involved in a broad range of physiological functions.[43]
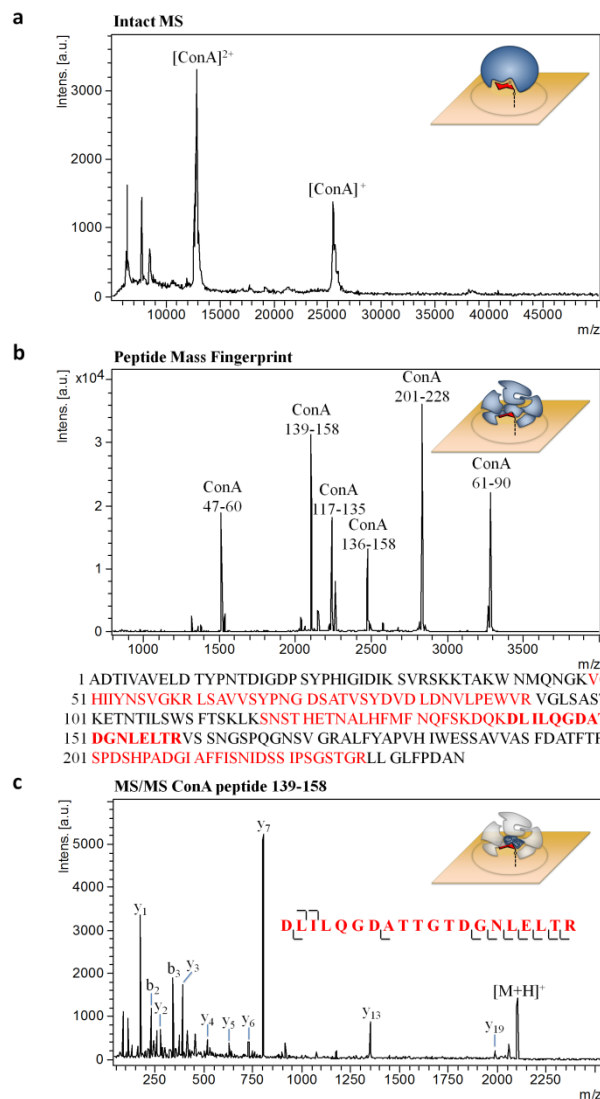
4

**Figure. 2** MALDI-ToF mass spectra of (a) ConA after incubation with a mannose-functionalized surface and (b) peptides generated after on-chip tryptic proteolysis (peptide mass fingerprint). Peptides identified are highlighted in red on the protein sequence. (c) The tandem mass spectrum of peptide 139-158 (*m/z* 2103.05), determined to be DLILQGDATTGTDGNLELTR, was unique to ConA in the Swiss-Prot database.

The interaction between anti-thrombin and heparin, for example, is dependent on a unique sulfated pentasaccharide sequence that is responsible for the potent anticoagulant properties of this widely used drug.[44] Thus, identifying proteins that interact with heparin / HS oligosaccharides could uncover novel biomedical roles for these complex polysaccharides.

Additional functionalization chemistries were required to immobilize these complex highly anionic structures onto the gold platform. The first was composed of heparin non-specifically and covalently immobilized to a polyethyleneimine (PEI) coated SAM, a surface that has been
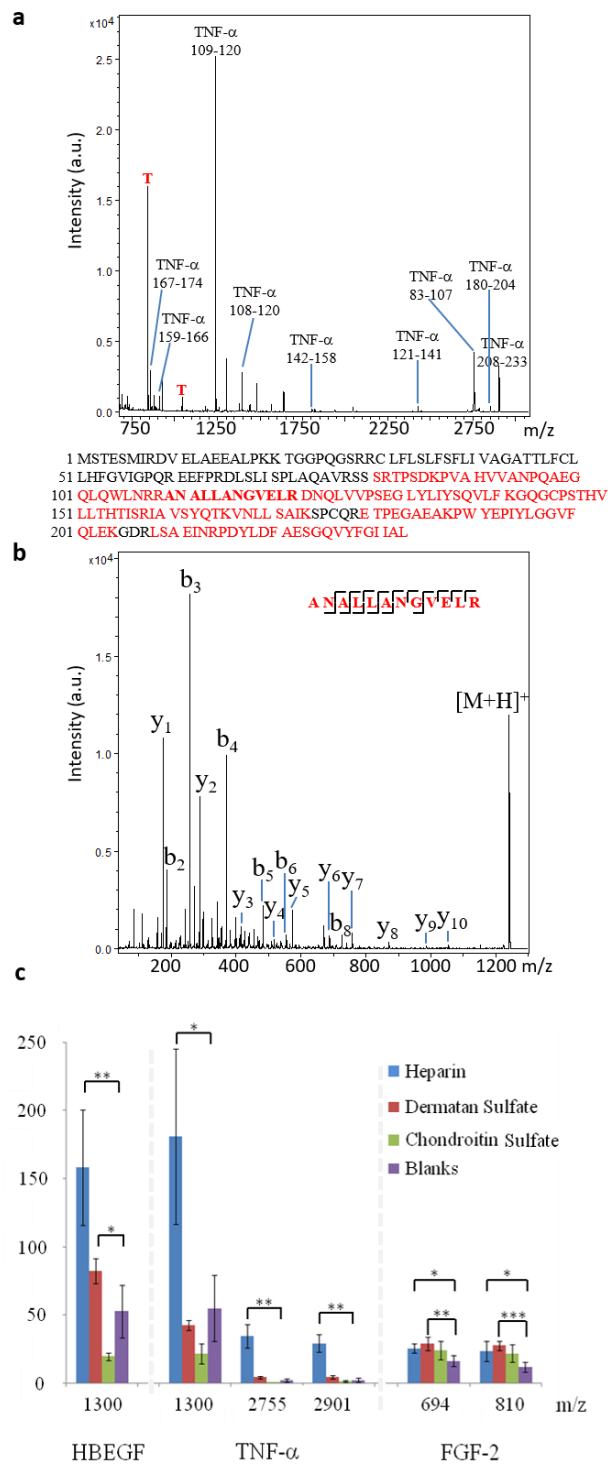


**Figure. 3** Peptide mass fingerprint after on-chip tryptic proteolysis of the heparin-binding protein TNF-α (human) after enrichment on the heparin coated surface (a). T corresponds to trypsin autolysis peptides. Matching peptides in the TNF-α sequence are highlighted in red. The ion at *m/z* 1240.8 was selected for tandem mass spectrometry and was identified as

5

ANALLANGVELR, a sequence unique to TNF-α in Swiss-Prot (b). The relative intensity of selected peptide ions derived from a series of known heparin-binding proteins (human) compared to the trypsin autolysis peak at $m/z$ 842 (n=4) is plotted after enrichment on various glycosaminoglycan arrays (c). Blank is composed of carboxymethyldextran (CMD). The relative intensity of the peptides on each GAG array was compared to the blank using a Student's t-test to determine statistical significance - * $p<0.05$, ** $p<0.01$ and *** $p<0.001$.

reported to minimize non-specific interactions with proteins without diminishing the affinity for GAG binding proteins.[45] Initially, 100 μg/mL of the known human heparin-binding proteins; tumor necrosis factor (TNF-α), heparin-binding EGF-like growth factor (HBEGF), and basic fibroblast growth factor (FGF2), were incubated on the heparin surface. Subsequent, tryptic digestion followed by MALDI-ToF MS/MS unequivocally identified the GAG binding proteins with >30% sequence coverage on all heparin-coated arrays (Figure. 3 and S8). Intensities of peptides derived from the GAG binding protein, relative to the trypsin autolysis peak ($m/z$ 842) were compared to that of the blank to ascertain relative binding. This minimized spot-to-spot and 'hot-spot' variability that can be observed in MALDI-ToF MS analysis). Incubation of these proteins with other GAG arrays reveals that FGF2 appears to demonstrate a weak affinity towards heparin, DS and CS in contrast to HBEGF and TNF-α that show a strong preference for heparin (Figure 3c). A second approach for GAGs is covalent coupling of smaller underivatized heparin oligosaccharides to SAMs via carbohydrazide chemistry.[26, 27] These arrays were used to explore the structural characteristics required for protein-saccharide interactions. Immobilized heparin oligosaccharides that differed in sulfation levels and numbers of disaccharide units from a single disaccharide (dp2) up to full length oligosaccharide (dp26) were incubated with FGF2 (Figure 4 and Figure S9). As a control, FGF2 was incubated with a heparin disaccharide since this structure is known to be too small to support FGF2 binding (Figure 4a).[46]

Tryptic peptides characteristic of FGF2 were observed for the tetrasaccharide (dp4) (Figure 4b) but not dp2, consistent with previous studies that indicate that the minimum binding length for heparin saccharides to FGF2 is a dp4.[46] Larger saccharides, such as dp26, also yielded FGF2 peptide signals (Figure 4c). To prove that FGF2 enrichment is the result of specificity and not the result of nonspecific ionic binding to the highly anionic surface, a nonheparin binding protein, Con A, was also tested. We observed no binding of Con A to any of the heparin oligosaccharides under physiological salt conditions (Figure S9). Overall, these studies confirm that the platform is suitable for screening GAG oligosaccharides for their protein binding properties. The studies also indicate that the platform may be used with a diverse range of functionali-

zation chemistries linking defined synthetic glycans, or polysaccharides/oligosaccharides from natural sources.
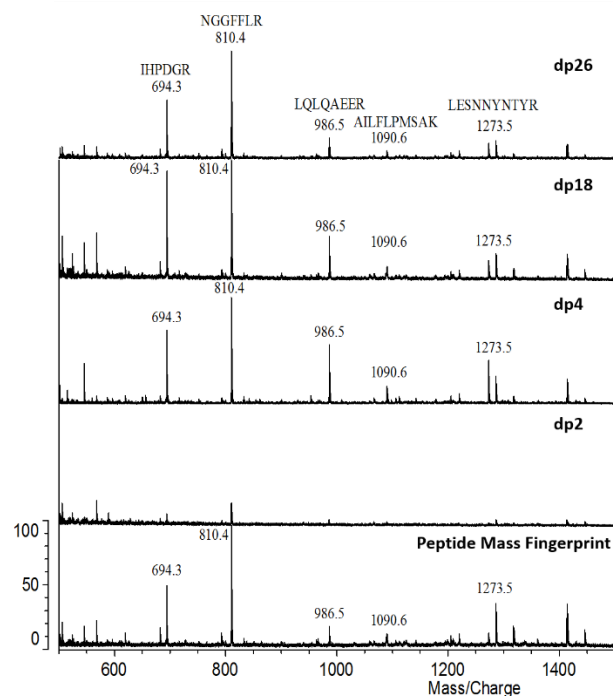


**Figure 4.** MALDI mass spectrometry of on-chip trypsin digested products following enrichment of FGF-2 on dp 2, 4, 18 and 26 heparin derivatized oligosaccharides. The peptide mass fingerprint of FGF2 is also shown for comparison.

## Lectin Arrays

Our final version of the array platform reversed the position of the binding partners by displaying a series of lectins on the surface and specifically enriching glycoproteins based on their glycan composition. A series of lectins were chosen that displayed orthogonal specificities for glycans indicative of both *O*- and *N*-glycan structures. Lectin arrays on other platforms such as polymers[9], glass slides[47] and metal surfaces[20] have already been shown to be valuable tools for the characterization of protein glycosylation. Lectin arrays have great potential in glycobiomarker discovery and as analytical platforms for the bioprocessing of biopharmaceuticals, many of which are glycoproteins.

Lectins were non-specifically and covalently immobilized to an EDC/NHS activated SAM. Ovalbumin (*Gallus gallus*) – a glycoprotein that contains N-glycans terminating in Man, GlcNAc and low amounts of Gal[48] (PNGase F digestion shown in Figure S13), and BSA – a non-glycosylated protein, were incubated with the lectins, and subsequently proteolyzed on-chip and analyzed by MALDI-ToF MS(/MS) (Figures S10 and S11). The intensities of the peptide ions which were derived from ovalbumin relative to the trypsin autolysis peak (n=3), were plotted in a heat map enabling rapid visualization of positive

6

binding events (Figure 5). As can clearly be seen from the heat map, ovalbumin is preferentially captured onto the Man and GlcNAc specific lectins ConA and WGA respectively, as would be expected. We were able to detect ovalbumin binding at concentrations of ~50μg/mL (0.7pmol load) (Figure S12), not dissimilar to concentrations employed on other array formats.[33, 49] We are currently attempting to lower the limit of detection by exploiting advances in MS technology (for example coupling of gas-phase separation techniques like ion-mobility spectrometry),[50] to approach those of fluorescence experiments without the requirement of potentially challenging chemical labeling.[10]
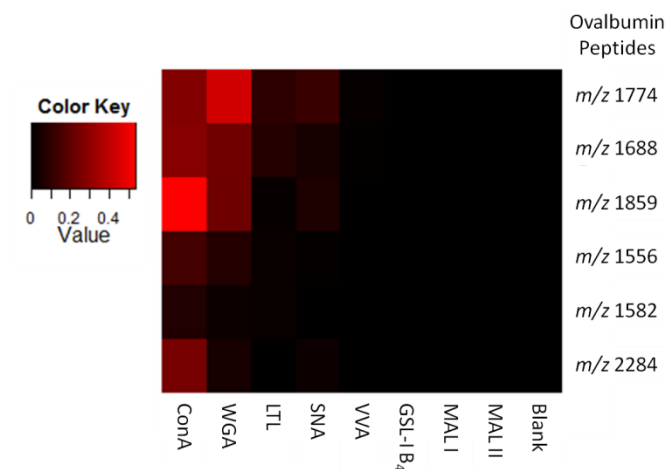


**Figure. 4** Heat map (n=3) displaying the intensities of ovalbumin-derived peptide ions normalized to the trypsin autolysis peak at $m/z$ 842, for each of the arrayed lectins. Blank corresponds to a self-assembled monolayer composed entirely of 'spacers'.

When the lectin array was probed with BSA control, no binding was observed as expected (Figure S11). Interestingly, precursor ions corresponding to the underlying lectins attached to the array surfaces were only really observed for ConA and SNA. Lack of signal could be due to low proteolytic efficiency due to steric protection, or in the case of WGA, its extensive disulfide bonding network. It is important to note that for the lectin arrays we have studied, there is little to no signal in the intact mass spectrum for either the enriched glycoprotein or the immobilized lectin. This is presumably due to the fact the proteins are strongly associated with the surface decreasing the likelihood of gas-phase ion formation.

**On-array affinity capture of human milk glycoproteins and GBPs**

To explore the potential of our arrays for enriching endogenous GBPs and glycoproteins from biofluids, a doubly defatted human milk sample was applied to the glycan, GAG and the lectin arrays. Human milk is well known to possess diverse glycoproteins and also heparin

binding proteins.[51, 52] Furthermore, understanding the milk glycome could be beneficial for therapeutic intervention given its importance in regulating the innate immune response of neonates. GeLC-MS was performed on the milk protein extract, with eight fractions being excised from the SDS gel (Figure 6a)[53] and subjected to in-gel trypsin digestion and subsequent nLC-MS/MS analysis (Figure S14). These initial studies confirmed the presence of several abundant proteins within the milk protein extract including: lactotransferrin (also known as lactoferrin), serum albumin, α-lactalbumin, Ig-α, β-casein and α-casein (Figure 6a and 6b, see electronic supplementary information section 6.1 for methodology).
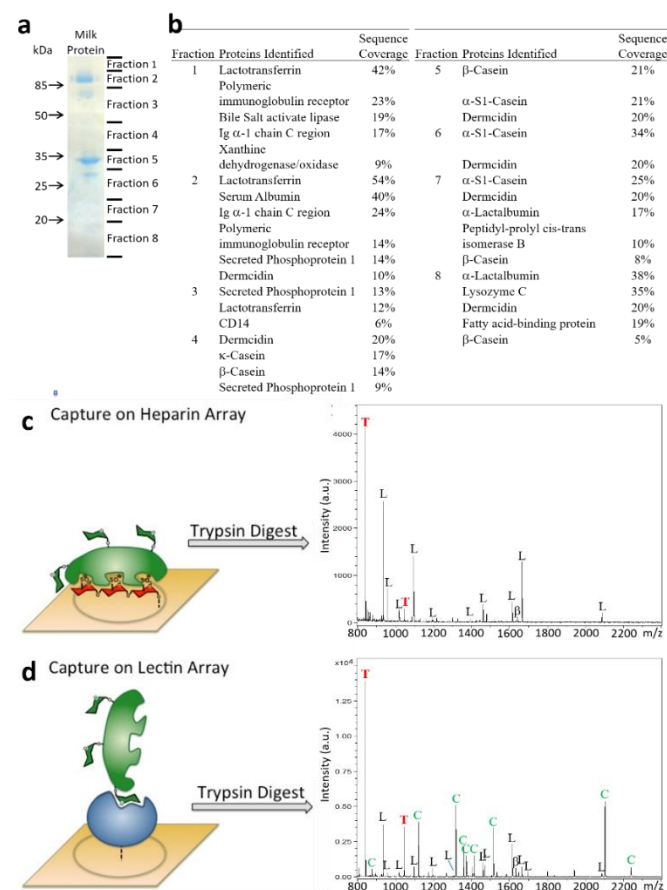


**Figure. 5** (a) SDS-PAGE of proteins extracted from human milk. (b) Proteins identified and proportional sequence coverage in each of the 8 fractions following in-gel trypsin digestion and nLC-MS/MS. (c) MALDI-ToF mass spectra of tryptic peptides after protein enrichment on a heparin array and (d) after incubation on a lectin functionalized array. B is a β-casein peptide; T are trypsin autolysis peptides, L are lactotransferrin peptides and C are ConA peptides. Table S1 details the peak identifications for milk incubated on ConA in greater detail.

Next, the milk protein extract was analyzed using the glycan, GAG, lectin arrays and without enrichment and after proteolytic digestion with trypsin. Peptides were identi-

7

fied by typical bottom-up proteomic strategies namely screening mass and tandem mass lists against protein databases (Swissprot and uniprot). When the milk protein is not enriched, only peptides derived from non-glycosylated β-Casein, α-Casein and κ-Casein are observed. After enrichment on the heparin arrays, only lactotransferrin was observed (Figure 6c). This is in agreement with previously reported studies.[51] Similarly, on the lectin arrays, the predominant peptides observed corresponded to lactotransferrin (Figure 6d, Table S1), which agrees with previous reports that this is one of the most abundant glycoproteins present within human milk.[52] Lactotransferrin possesses both complex and hybrid *N*-glycans terminating in β-Gal, β-Glc*N*Ac, α-Neu5Ac, α-Man and α-Fuc,[54] explaining why it is enriched by multiple lectins on the arrays, including; ConA, LTL, SNA, GSL-1 B$_4$ and MAL I (Figures S15 and S16). Enrichment of lactotransferrin by LTL suggests that the lectin has a very low or no affinity for a single Fuc residue and requires a greater amount of the glycan structure i.e. the Lewis x moiety. A single β-casein peptide (VLPIPQQVVPYPQR, *m/z* 1634) can also be observed in most of the spectra (Figures 6b, S15 and S16); β-casein has not been reported to be glycosylated so this is presumably due to minimal non-specific binding. No other peptides derived from β-casein were observed in the mass spectra suggesting that low level non-specifically bound β-casein may yield this peptide either due to rapid proteolysis and/or relatively high ionization efficiency. Indeed, the MALDI-ToF spectra of crude protein extracts from human milk used for these studies were dominated by peptide ions derived from casein (Figure S15a), demonstrating the considerable enrichment of binding proteins on the array platform. Thus, we have demonstrated that we can unambiguously identify both glycan-binding proteins and glycoproteins among the high abundant proteins from a biofluid with the array platforms presented here, which to the best of our knowledge has not been achieved in the literature.

## Conclusions

In summary, we present a versatile and robust glycomics array platform. The platform enables the enrichment and identification of glycan-protein interactions from complex biological extracts without the requirement of a chemical label or pre-purification of samples. Three different surface functionalizations are described. We have shown that the arrays can be prepared with a range of defined chemically synthesized mono-, di- and tri-saccharides. More complex polysaccharides, in particular glycosaminoglycans which are important component of the cell matrix, are also fully compatible. The flexibility to use both of synthetic and natural glycan probes allows us to capture a broad range of cell-surface glycans, neither being limited by chemical synthesis of natural availability. Additionally, we have shown that the array surface can be

functionalized with carbohydrate-binding proteins (generating lectin arrays) for interrogation of glycoconjugates. We have shown that both the immobilized glycans and lectins retain their ability to bind GBPs and glycoproteins respectively, in line with published data on other platforms.

The platform was applied to investigate a number of biological samples: a lectin (wheat germ agglutinin) was captured from a crude wheat germ protein extract. Investigation of unfractionated human milk proteins with either heparin arrays or lectin arrays allowed for the capture and identification of lactotransferrin as a heparin-binding glycoprotein. Critically, high-throughput label-free detection and characterization of these diverse specifically bound proteins was achieved using a combination of MALDI-ToF MS/MS, on-chip proteolysis and Mascot-driven database searches. Although the focus of this article was limited to studying mainly glycans and their associated protein binding partners, our array technology could be readily applied to other biologically relevant molecules. The results presented highlight the future potential of our platform in discovering unknown glycan binding partners and for the rapid screening of clinical biomarkers possessing aberrant glycosylation patterns.

## ASSOCIATED CONTENT

**Supporting Information**.
Selected mass spectra, additional methodologies and synthesis are available *via* the Internet.

## AUTHOR INFORMATION

### Corresponding Author

* E-mail: Sabine.Flitsch@Manchester.ac.uk, Claire.Eyers@Liverpool.ac.uk, J.Turnbull@Liverpool.ac.uk

### Present Addresses

§ Chemie-Cluster Bayern GmbH, Germany
ϕ *Syngenta, UK*

### Author Contributions

∥These authors contributed equally.

## REFERENCES

(1) Varki A, C. R., Esko JD, et al. *Essentials of Glycobiology*, 2 ed.; Cold Spring Harbor: New York, 2009.

(2) Disney, M. D.; Seeberger, P. H. *Chemistry & Biology* **2004**, *11*, 1701-1707.

8

(3)   Adams, E. W.; Ratner, D. M.; Bokesch, H. R.; McMahon, J. B.; O'Keefe, B. R.; Seeberger, P. H. *Chemistry & Biology* **2004**, *11*, 875-881.

(4)   Ratner, D. M.; Adams, E. W.; Su, J.; O'Keefe, B. R.; Mrksich, M.; Seeberger, P. H. *Chembiochem* **2004**, *5*, 379-382.

(5)   Hakomori, S. *Cancer Research* **1996**, *56*, 5309-5318.

(6)   Gornik, O.; Gornik, I.; Gasparovic, V.; Lauc, G. *Clinical Biochemistry* **2008**, *41*, 504-510.

(7)   Yoshida-Moriguchi, T.; Yu, L. P.; Stalnaker, S. H.; Davis, S.; Kunz, S.; Madson, M.; Oldstone, M. B. A.; Schachter, H.; Wells, L.; Campbell, K. P. *Science* **2010**, *327*, 88-92.

(8)   Seeberger, P. H.; Werz, D. B. *Nature* **2007**, *446*, 1046-1051.

(9)   Pilobello, K. T.; Krishnamoorthy, L.; Slawek, D.; Mahal, L. K. *Chembiochem* **2005**, *6*, 985-989.

(10)  Kuno, A.; Uchiyama, N.; Koseki-Kuno, S.; Ebe, Y.; Takashima, S.; Yamada, M.; Hirabayashi, J. *Nature Methods* **2005**, *2*, 851-856.

(11)  Zheng, T.; Peelen, D.; Smith, L. M. *Journal of the American Chemical Society* **2005**, *127*, 9982-9983.

(12)  Laurent, N.; Voglmeir, J.; Wright, A.; Blackburn, J.; Pham, N. T.; Wong, S. C. C.; Gaskell, S. J.; Flitsch, S. L. *Chembiochem* **2008**, *9*, 883-887.

(13)  Voglmeir, J.; Sardzik, R.; Weissenborn, M. J.; Flitsch, S. L. *Omics-a Journal of Integrative Biology* **2010**, *14*, 437-444.

(14)  Sardzik, R.; Sharma, R.; Kaloo, S.; Voglmeir, J.; Crocker, P. R.; Flitsch, S. L. *Chemical Communications* **2011**, *47*, 5425-5427.

(15)  Sardzik, R.; Green, A. P.; Laurent, N.; Both, P.; Fontana, C.; Voglmeir, J.; Weissenborn, M. J.; Haddoub, R.; Grassi, P.; Haslam, S. M.; Widmalm, G.; Flitsch, S. L. *Journal of the American Chemical Society* **2012**, *134*, 4521-4524.

(16)  Both, P.; Green, A. P.; Gray, C. J.; Šardzík, R.; Voglmeir, J.; Fontana, C.; Austeri, M.; Rejzek, M.; Richardson, D.; Field, R. A.; Widmalm, G.; Flitsch, S. L.; Eyers, C. E. *Nat Chem* **2014**, *6*, 65-74.

(17)  Sanchez-Ruiz, A.; Serna, S.; Ruiz, N.; Martin-Lomas, M.; Reichardt, N. C. *Angewandte Chemie-International Edition* **2011**, *50*, 1801-1804.

(18)  Ban, L.; Pettit, N.; Li, L.; Stuparu, A. D.; Cai, L.; Chen, W.; Guan, W.; Han, W.; Wang, P. G.; Mrksich, M. *Nature Chemical Biology* **2012**, *8*, 769-773.

(19)  Serna, S.; Yan, S.; Martin-Lomas, M.; Wilson, I. B. H.; Reichardt, N. C. *Journal of the American Chemical Society* **2011**, *133*, 16495-16502.

(20)  Morris, T. A.; Peterson, A. W.; Tarlov, M. J. *Analytical Chemistry* **2009**, *81*, 5413-5420.

(21)  Ueda, K.; Fukase, Y.; Katagiri, T.; Ishikawa, N.; Irie, S.; Sato, T. A.; Ito, H.; Nakayama, H.; Miyagi, Y.; Tsuchiya, E.; Kohno, N.; Shiwa, M.; Nakamura, Y.; Daigo, Y. *Proteomics* **2009**, *9*, 2182-2192.

(22)  Beloqui, A.; Calvo, J.; Serna, S.; Yan, S.; Wilson, I. B. H.; Martin-Lomas, M.; Reichardt, N. C. *Angewandte Chemie International Edition* **2013**, *52*, 7477-7481.

(23)  Royle, L.; Roos, A.; Harvey, D. J.; Wormald, M. R.; Van Gijlswijk-Janssen, D.; Redwan, E.-R. M.; Wilson, I. A.; Daha, M. R.; Dwek, R. A.; Rudd, P. M. *Journal of Biological Chemistry* **2003**, *278*, 20140-20153.

(24)  Newburg, D. S. *Journal of Animal Science* **2009**, *87*, 26-34.

(25)  Love, J. C.; Estroff, L. A.; Kriebel, J. K.; Nuzzo, R. G.; Whitesides, G. M. *Chemical Reviews* **2005**, *105*, 1103-1169.

(26)  Zhi, Z.-L.; Laurent, N.; Powell, A. K.; Karamanska, R.; Fais, M.; Voglmeir, J.; Wright, A.; Blackburn, J. M.; Crocker, P. R.; Russell, D. A.; Flitsch, S.; Field, R. A.; Turnbull, J. E. *Chembiochem* **2008**, *9*, 1568-1575.

(27)  Zhi, Z.-l.; Powell, A. K.; Turnbull, J. E. *Analytical Chemistry* **2006**, *78*, 4786-4793.

9

(28) Grant, O. C.; Smith, H. M.; Firsova, D.; Fadda, E.; Woods, R. J. *Glycobiology* **2014**, *24*, 17-25.

(29) Sardzik, R.; Noble, G. T.; Weissenborn, M. J.; Martin, A.; Webb, S. J.; Flitsch, S. L. *Beilstein Journal of Organic Chemistry* **2010**, *6*, 699-703.

(30) Powell, A. K.; Ahmed, Y. A.; Yates, E. A.; Turnbull, J. E. *Nature Protocols* **2010**, *5*, 821-833.

(31) Dallas, D. C.; Martin, W. F.; Strum, J. S.; Zivkovic, A. M.; Smilowitz, J. T.; Underwood, M. A.; Affolter, M.; Lebrilla, C. B.; German, J. B. *Journal of Agricultural and Food Chemistry* **2011**, *59*, 4255-4263.

(32) Blixt, O.; Head, S.; Mondala, T.; Scanlan, C.; Huflejt, M. E.; Alvarez, R.; Bryan, M. C.; Fazio, F.; Calarese, D.; Stevens, J.; Razi, N.; Stevens, D. J.; Skehel, J. J.; van Die, I.; Burton, D. R.; Wilson, I. A.; Cummings, R.; Bovin, N.; Wong, C.-H.; Paulson, J. C. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 17033-17038.

(33) Song, X.; Xia, B.; Stowell, S. R.; Lasanajak, Y.; Smith, D. F.; Cummings, R. D. *Chemistry & Biology* **2009**, *16*, 36-47.

(34) Ni, J. H.; Singh, S.; Wang, L. X. *Bioconjugate Chemistry* **2003**, *14*, 232-238.

(35) Kumar, S. R.; Sauter, E. R.; Quinn, T. P.; Deutscher, S. L. *Clinical Cancer Research* **2005**, *11*, 6868-6871.

(36) Debray, H.; Decout, D.; Strecker, G.; Spik, G.; Montreuil, J. *European Journal of Biochemistry* **1981**, *117*, 41-55.

(37) Iskratsch, T.; Braun, A.; Paschinger, K.; Wilson, I. B. H. *Analytical Biochemistry* **2009**, *386*, 133-146.

(38) Geisler, C.; Jarvis, D. L. *Glycobiology* **2011**, *21*, 988-993.

(39) Porter, A.; Yue, T.; Heeringa, L.; Day, S.; Suh, E.; Haab, B. B. *Glycobiology* **2010**, *20*, 369-380.

(40) Yan, L. Y.; Wilkins, P. P.; AlvarezManilla, G.; Do, S. I.; Smith, D. F.; Cummings, R. D. *Glycoconjugate Journal* **1997**, *14*, 45-55.

(41) Cottrell, J. S. *Peptide Research* **1994**, *7*, 115-&.

(42) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551-3567.

(43) Bishop, J. R.; Schuksz, M.; Esko, J. D. *Nature* **2007**, *446*, 1030-1037.

(44) Bjork, I.; Lindahl, U. *Molecular and Cellular Biochemistry* **1982**, *48*, 161-182.

(45) Pathak, A.; Kumar, P.; Chuttani, K.; Jain, S.; Mishra, A. K.; Vyas, S. P.; Gupta, K. C. *ACS Nano* **2009**, *3*, 1493-1505.

(46) Delehedde, M.; Lyon, M.; Gallagher, J. T.; Rudland, P. S.; Fernig, D. G. *Biochemical Journal* **2002**, *366*, 235-244.

(47) Li, C.; Simeone, D. M.; Brenner, D. E.; Anderson, M. A.; Shedden, K. A.; Ruffin, M. T.; Lubman, D. M. *Journal of Proteome Research* **2009**, *8*, 483-492.

(48) Lattova, E.; Perreault, H. N.; Krokhin, O. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 725-735.

(49) Etxebarria, J.; Calvo, J.; Martin-Lomas, M.; Reichardt, N. C. *Acs Chemical Biology* **2012**, *7*, 1729-1737.

(50) Baker, E. S.; Burnum-Johnson, K. E.; Ibrahim, Y. M.; Orton, D. J.; Monroe, M. E.; Kelly, R. T.; Moore, R. J.; Zhang, X.; Théberge, R.; Costello, C. E.; Smith, R. D. *Proteomics* **2015**, n/a-n/a.

(51) Zou, S.; Magura, C. E.; Hurley, W. L. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **1992**, *103*, 889-895.

(52) Lönnerdal, B. *The American Journal of Clinical Nutrition* **1985**, *42*, 1299-1317.

(53) Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M. *Nat. Protocols* **2007**, *1*, 2856-2860.

(54) Barboza, M.; Pinzon, J.; Wickramasinghe, S.; Froehlich, J. W.; Moeller, I.; Smilowitz, J. T.; Ruhaak, L. R.; Huang, J.; Lönnerdal, B.; German, J. B.; Medrano, J. F.; Weimer, B. C.; Lebrilla, C. B. *Molecular & Cellular Proteomics* **2012**, *11*.

10