



# O-Glycome Beam Search Arrays for Carbohydrate Ligand Discovery\*<sup>§</sup>

Zhen Li<sup>‡‡</sup>, Chao Gao<sup>‡‡</sup>, Yibing Zhang<sup>‡</sup>, Angelina S. Palma<sup>‡¶</sup>, Robert A. Childs<sup>‡</sup>, Lisete M. Silva<sup>‡</sup>, Yang Liu<sup>||</sup>, Xi Jiang<sup>||\*\*</sup>, Yan Liu<sup>‡</sup>, Wengang Chai<sup>‡§§</sup>, and Ten Feizi<sup>‡§§</sup>

**O-glycosylation is a post-translational modification of proteins crucial to molecular mechanisms in health and disease. O-glycans are typically highly heterogeneous. The involvement of specific O-glycan sequences in many bio-recognition systems is yet to be determined because of a lack of efficient methodologies. We describe here a targeted microarray approach: O-glycome beam search that is both robust and efficient for O-glycan ligand-discovery. Substantial simplification of the complex O-glycome profile and facile chromatographic resolution is achieved by arraying O-glycans as branches, monitoring by mass spectrometry, focusing on promising fractions, and on-array immuno-sequencing. This is orders of magnitude more sensitive than traditional methods. We have applied beam search approach to porcine stomach mucin and identified extremely minor components previously undetected within the O-glycome of this mucin that are ligands for the adhesive proteins of two rotaviruses. The approach is applicable to O-glycome recognition studies in a wide range of biological settings to give insights into glycan recognition structures in natural microenvironments. *Molecular & Cellular Proteomics 17: 10.1074/mcp.RA117.000285, 121–133, 2018.***

Mucins are glycoproteins crucial in molecular mechanisms in health and disease (1–3). They occur in secretions and in membrane-associated forms on epithelia, endothelia and leu-

kocytes being associated with tissue- and cell-specific activities such as cell-attachment of commensal and pathogenic microbes, targeting of leukocytes to endothelial cells and modulation of immune responses (4). The glycans on mucins are O-glycosidically linked to serine or threonine, and range from a single to 20 or more monosaccharide residues (5). The O-glycans are known to express species- and tissue-specific antigens that change during cell differentiation and oncogenesis (2, 6), and mediate some of the above-mentioned mucin functions (7). However, the involvements of specific O-glycan structures in many processes are yet to be determined (2), assignments being hampered by the diversity of extensions at the core-monosaccharide, GalNAc, and multiplicity of chain lengths and branching patterns (5, 8, 9), and often the limited amounts of glycans available. Thus high-throughput and microscale methodologies are needed for deconvolution and functional assignments of individual O-glycans within O-glycomes as cues to large scale syntheses of desired glycan sequences for exploitation (10, 11).

In classical studies, O-glycans released from abundantly available mucins (from ovarian cystadenomas) were used to elucidate the structures of the major blood group antigens A, B, H and Lewis<sup>a</sup> (Le<sup>a</sup>) and Le<sup>b</sup> (12, 13). O-glycans or their fragments were used as inhibitors of antibody binding requiring large amounts (milligrams) of individually purified glycans. To address the need for a microscale method for determining recognition of individual glycans by antibodies and other carbohydrate-binding proteins, the neoglycolipid (NGL)<sup>1</sup> technology was introduced whereby glycans, singly or as mixtures, are conjugated to a lipid molecule, and immobilized on solid matrices, resolved by thin layer chromatography (TLC) and probed with carbohydrate-binding proteins (14–16). The NGL

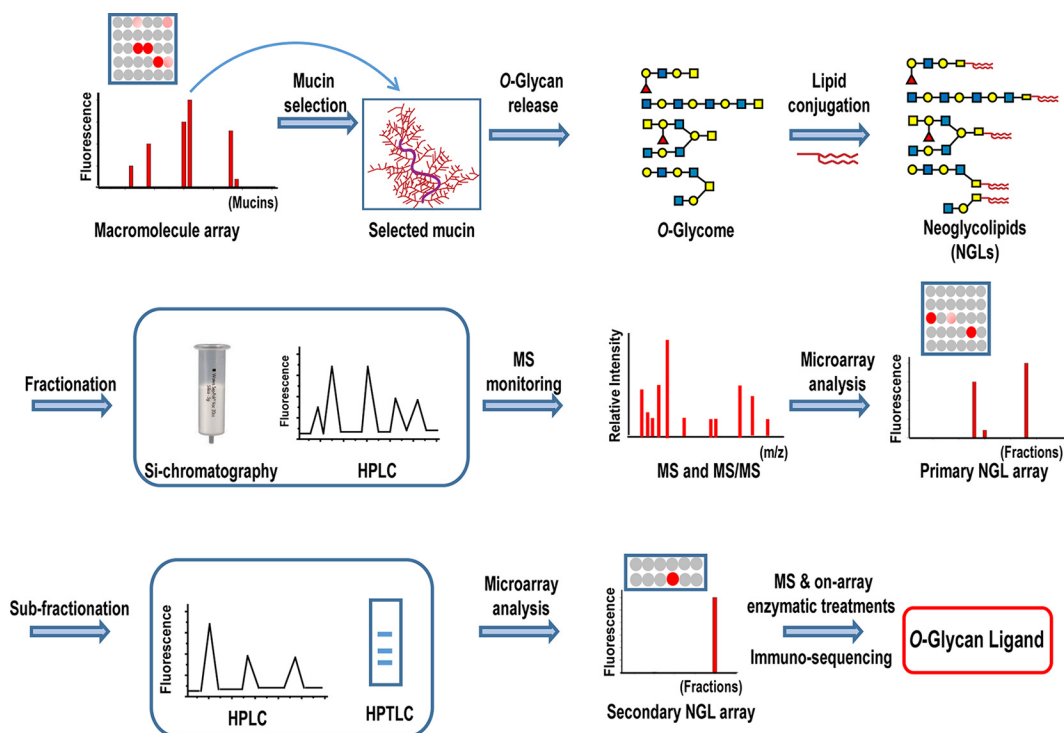
From the <sup>‡</sup>Glycosciences Laboratory, Department of Medicine, Imperial College London, W12 0NN, UK; <sup>§</sup>Department of Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215; <sup>¶</sup>Department of Chemistry, UCIBIO-NOVA University of Lisbon, 1099085, Portugal; <sup>||</sup>Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center and <sup>\*\*</sup>University of Cincinnati College of Medicine, Cincinnati, Ohio 45229

✂ Author's Choice—Final version free via Creative Commons CC-BY license.

Received August 22, 2017, and in revised form, November 13, 2017  
Published, MCP Papers in Press, November 28, 2017, DOI 10.1074/mcp.RA117.000285

Author contributions. T.F., Yan Liu, and W.C. designed research. Z.L. and C.G. designed O-glycome arrays and performed analyses. Yang Liu and X.J. contributed proteins. A.S.P. generated mucin arrays; C.G. generated the F77/li array and designed on-array fucosidase experiments; R.A.C. and L.M.S. contributed to analyses. Y.Z. performed mass spectrometry. Z.L., T.F., W.C., C.G., and Yan Liu wrote the paper and A.S.P., L.M.S., and R.A.C. contributed to writing. All authors commented on the paper.

<sup>1</sup> The abbreviations used are: NGL, neoglycolipid; ADHP, N-ami-noacetyl-N-(9-anthracenylmethyl)-1,2-dihexadecyl-sn-glycero-3-phosphoethanolamine; BSM, bovine submaxillary mucin; DHPE, 1,2-dihexadecyl-sn-glycero-3-phosphoethanolamine; ESI-MS, electro-spray-ionization induced MS; ESI-CID-MS/MS, electro spray ionization collision induced MS/MS; GST, glutathione S-transferase; HPTLC, high performance thin layer chromatography; LNFPI, lacto-N-fucopentaose I; LNnFPI, lacto-N-neofucopentaose I; LNDfH-I, Lacto-N-difucohexaose; LNFPII, Lacto-N-fucopentaose II; MS, mass spectrometry; MALDI-MS, matrix assisted laser desorption/ionization; PSM, porcine stomach mucin; pLNH, p-Lacto-N-hexaose; pLNnH, p-Lacto-N-neohexaose; UEA-I, Ulex europaeus agglutinin-I.



**FIG. 1. Scheme for the O-glycome beam search approach to target naturally occurring ligands within heterogeneous populations.** The macromolecule array is of mucins. A ligand-bearing mucin is selected after microarray analyses of a carbohydrate recognizing protein. O-glycan alditols released from the mucin are converted, after mild periodate oxidation, to fluorescent neoglycolipids (NGLs) derived from branches 3- or 6-linked to core *N*-acetylgalactosaminitol with resulting marked simplification of glycan profiles. The NGLs are fractionated by chromatographies such as silica chromatography and HPLC, monitored by mass spectrometry (MS) and robotically arrayed. These fractions constitute the primary array. Microarray analyses using the primary array enable targeting of ligand-positive fractions. These are sub-fractionated as necessary with iterative secondary or tertiary microarray analyses to identify, isolate and characterize the ligands by MS and various microscale structural analyses including on-array enzymatic treatments and immuno-sequencing. Symbol Nomenclature for Glycans (SNFG) is used for monosaccharides: ●, galactose; ■, *N*-acetylglucosamine; ▲, fucose; ◻, *N*-acetylgalactosamine. For simplicity the modification of the core *N*-acetylgalactosaminitol is not shown.

technology coupled with mass spectrometry has undergone sequential developments (17, 18), and has been the basis of the first microarray system for sequence-defined glycans (19). This is now a state-of-the-art platform with the glycan probes robotically arrayed in a liposomal formulation at low femtomoles per spot (20, 21). A “designer” array of NGLs from an epithelial O-glycome, resolved on TLCs enabled the characterization of the elusive prostate cancer-associated antigen F77 (22).

The set of microscale methods described here for O-glycan populations (Fig. 1) is a significant advance involving, as an early step, microarray analysis of NGL populations resolved by HPLC with mass spectrometry (MS) monitoring rather than TLC plate binding analyses (commonly referred to as chromatogram binding), which necessitates elution of bound components (22). There follows iterative microarray analyses of targeted ligand-positive fractions resolved by appropriate chromatographic separations concomitantly with MS and culminating in micro-immuno-sequencing of a ligand-positive glycan. We coin the term beam search because the power to search for and focus on ligand-positive components in a

population of hundreds of O-glycans within an epithelial O-glycome. As an exemplar study-case, applying the approach to a ligand-bearing mucin, we identify O-glycan ligands for the cell-adhesion proteins, VP8\*, of two rotaviruses P[19] and P[10]. These are evolutionarily closely related but of distinct genotypes. The first, infects humans and pigs, and the second has so far been reported to infect humans. Like VP8\* proteins of other rotaviruses, those of P[10] and P[19] bind human salivary mucins and porcine stomach mucin (PSM), but the glycan ligands in these have not been characterized (23, 24). The approach is complemented by microarrays of sequence-defined glycans.

#### EXPERIMENTAL PROCEDURES

**Recombinant VP8\* Proteins of P[10] and P[19] Rotaviruses**—The recombinant rotavirus VP8\* proteins were expressed as glutathione S-transferase (GST)-tagged proteins as described (25).

**Monoclonal Anti-blood group Antibodies (MAbs) and Plant Lectins**—Murine anti-A type 1 IgM (AH21), anti-A Lewis<sup>b</sup> IgG (HH3) were gifts from Henrik Clausen, Center for Glycomics, University of Copenhagen. Anti-A type 2 IgM (Z2A), anti-A IgG (T36), anti-H type 1 IgG (17–206), and anti-H type 2 IgG (BRIC231), were from Abcam

(Cambridge, UK). Anti-lacto-*N*-tetraose (LNT) IgM (SIG-3310) was from Signet Laboratories (Hayward, CA). Biotinylated lectins *Ulex europaeus* agglutinin I (UEA-I) and *Aleuria aurantia* lectin (AAL), were from Vector Laboratories (Peterborough, UK).

**Mucins and Mucin-enriched Glycoproteins**—Porcine stomach mucin (PSM) was from Sigma (Dorset, UK); other mucins and mucin-enriched glycoproteins in [supplemental Table S1](#) have been described previously (22).

**O-Glycans**—Steps in the release and processing of PSM O-glycome and derived NGL fractions for robotic arraying are in [supplemental Fig. S1](#). O-Glycans were released from PSM, 1 g dry weight, by reductive alkaline borohydride degradation (22) and fractionated initially by anion-exchange chromatography on an AG1X8 column (Bio-Rad, Hercules, CA, acetate form, 3 ml resin), with sequential elution with 2 mM and 300 mM pyridium acetate, followed by 300 mM Na<sub>2</sub>SO<sub>4</sub> to yield predominantly neutral, sialylated and sulfated fractions, respectively. The neutral fraction was the subject of the present study. The neutral fraction was chromatographed on a Bio-Gel P4 (Bio-Rad) column (1.6 × 90 cm) eluted with water at a flow rate of 15 ml/h. Fractions were monitored on-line by refractive index and their carbohydrate contents with respect to hexose were determined by the dot orcinol-sulfuric acid method using galactose as standard (18).

**NGLs**—Fluorescent NGLs were prepared from the reductively released oligosaccharides (alditols) using *N*-aminoacetyl-*N*-(9-anthracenylmethyl)-1,2-dihexadecyl-sn-glycero-3-phosphoethanolamine (ADHP) after mild periodate oxidation (18). In brief, lyophilized O-glycan fractions from the Bio-Gel P4 column (~1 μmol of each, estimated based on the average molecular weights and the sugar contents in the fractions as in [supplemental Table S2](#)) were mixed with freshly prepared sodium periodate (700 μl, 1.25 mg/ml in imidazole buffer, pH 6.5). Reaction mixtures were kept on ice in the dark for 5 min followed by addition of 80 μl butane-2,3-diol (9 mg/ml) and incubation in the dark for a further 40 min. ADHP (2 ml of 10 mM solution in CHCl<sub>3</sub>/MeOH, 1:1 by volume) and 1.45 ml freshly prepared tetrabutylammonium cyanoborohydride (20 mg/ml in MeOH) were then added, and the reaction carried out at 60 °C for 16–24 h. Reaction mixtures were analyzed by HPTLC developed with CHCl<sub>3</sub>/MeOH/H<sub>2</sub>O, 60:35:8, and visualized under long-wave length UV light. NGLs were quantified with respect to their fluorescent lipid contents by spectrophotometry at 368.4 nm using ADHP as a standard (22).

A glycolipid and 11 NGLs (18, 20, 22) were included as reference compounds in the tertiary array designated “Fucase Array” (see below). Their sequences are given in [supplemental Table S6](#). These included H2, LNFP-I.DH (lacto-*N*-fucopentaose I), LNnFP-I.DH (lacto-*N*-neofucopentaose I), GSC-915–3.DH, GSC-915–4, pLNH.DH (*p*-Lacto-*N*-hexaose), pLNH-b.DH, pLNnH.DH (*p*-Lacto-*N*-neohexaose), LNDfH-I.DH (Lacto-*N*-difucohexaose I), LNFP-II.DH (Lacto-*N*-fucopentaose II) (22, 26).

**High-performance Liquid Chromatography (HPLC)**—NGLs were fractionated by normal phase HPLC with a XBridge Amide column (3.5 μm, 4.6 × 250 mm, Waters, Borehamwood, UK) using a Gilson system coupled with a fluorescence detector (λ<sub>ex</sub> at 255 nm and λ<sub>em</sub> at 405 nm) (22). The gradient was CHCl<sub>3</sub>/MeOH/H<sub>2</sub>O 130:70:9 (solvent A) to 10:20:8 (solvent B) in 60 min at a flow rate of 0.5 ml/min.

**Semipreparative High Performance Thin Layer Chromatography (HPTLC)**—For semi-preparative TLC, NGLs were dissolved in CHCl<sub>3</sub>/MeOH/H<sub>2</sub>O, 25:25:8 and applied on the aluminum backed HPTLC plates (Merck, Darmstadt, Germany), and were chromatographed in CHCl<sub>3</sub>/MeOH/H<sub>2</sub>O, 60:35:8. For improved separation, the plate was dried and developed in the same solvent for a second time. The NGLs were visualized under 254 nm UV light. The TLC bands of interest were scraped off the plates, and the NGLs eluted with CHCl<sub>3</sub>/MeOH/H<sub>2</sub>O, 25:25:8 from a mini-column packed with the silica gel.

**Mass Spectrometry**—Negative-ion electrospray-ionization MS (ESI-MS) and electrospray ionization collision-induced MS/MS (ESI-CID-MS/MS) analyses of oligosaccharides were carried out on a Waters Synapt G2-S instrument. Oligosaccharides were dissolved in water at 5–10 pmol/μl and 1–2 μl was loop-injected for analysis. ACN/H<sub>2</sub>O (1:1) containing 5 mM NH<sub>4</sub>HCO<sub>3</sub> was used as the elution solvent. Cone voltage 80 V and the collision gas Ar with collision energies of 15–30 V were used for CID-MS/MS. Matrix-assisted laser desorption ionization (MALDI-MS) and MALDI-MS/MS analyses of NGLs were carried out in both negative- and positive-ion mode on Shimadzu AXIMA Assurance (linear TOF) and Resonance (QIT-TOF) instrument, respectively. Oligosaccharides and NGLs were dissolved in water and CHCl<sub>3</sub>/MeOH/H<sub>2</sub>O 25:25:8, respectively, at a concentration of 10–20 pmol/μl, and 0.5 μl was deposited on the sample target together with 1 μl of matrix: 2',4',6'-trihydroxyacetophenone. Ethanol was used for recrystallization. For CID-MS/MS, a collision gas Ar (2 bar) and optimized collision energy in the region of 80–140 V were used to obtain fragmentation.

**Glycan Microarrays**—Five microarrays on 16-pad nitrocellulose-coated glass slides (Sartorius Stedim, Goettingen, Germany) using a noncontact arrayer, Nano-Plotter (Gesim, Germany) were used in the present study. These included (a) the macromolecule array of mucin-rich glycoproteins; (b) the primary NGL array of 60 fractions of the PSM neutral O-glycome resolved by HPLC ([supplemental Table S4](#)); (c) the secondary NGL array designated Fucase Array ([supplemental Table S6](#)) which was generated for on-array fucosidase digestion experiments and consisted of NGLs of four PSM O-glycan subfractions and 12 sequence-defined lipid-linked glycan probes; and (d) a screening array of sequence-defined glycans designated “F77/li array” consisting of 60 lipid-linked glycan probes ([supplemental Table S7](#)), in part described previously (19), which was used in the validation of the beam search results; and (e) a small array of fungus-derived gluco oligosaccharides ([supplemental Table S8](#)).

The macromolecule array was of 30 mucin-enriched epithelial glycoproteins designated “Mucin Array” (22), which consisted of extracts from ovarian cystadenoma fluids, meconia, several purified mucins of human origin and two of animal origins (PSM and BSM). Two array sets of these were used: mucin set 2 was used in the analyses of the VP8\* proteins in 2. 2A and 2B, and mucin set 3 in analyses of the antibodies and lectins in [supplemental Table S1](#) and [supplemental Fig. S2](#). These were arrayed in water (22) at 30 and 170 pg hexose per spot. Binding signals were glycoprotein dose-related; results at 170 pg hexose per spot level are shown.

The primary and secondary NGL arrays, and the “F77/li array” of lipid-linked glycan probes (NGLs and glycolipids) were generated in a liposomal formulation (20). In brief the printing solutions contained 100 pmol/μl of phosphatidylcholine and cholesterol (both from SIGMA) as lipid carriers in addition to the lipid-linked glycan probes in water (HPLC grade). The concentrations of the lipid-linked glycan probes were 5 and 15 pmol/μl for the 2 and 5 fmol per spot levels, respectively. The printing solutions also contained Cy3 NHS ester (GE Healthcare, Buckinghamshire, UK) at 20 ng/ml (26 fmol/μl) as a marker to monitor the printing process.

Microarray binding analyses were performed at ambient temperature, essentially as described (20). In brief, for analyses of the viral proteins, the arrayed slides were blocked for 1 h with 0.02% (w/v) casein (Pierce, ThermoFischer Scientific, Walham, MA) and 1% bovine serum albumin (BSA, Sigma) in HBS (5 mM HEPES buffer pH 7.4, 150 mM NaCl) with 5 mM CaCl<sub>2</sub> (Blocker A). The VP8\* proteins of P[10] and P[19] viruses were diluted to 25 μg/ml, 50 μg/ml or 100 μg/ml, as indicated, in Blocker A and overlaid for 1.5 h, followed detection for 1 h with rabbit anti-GST polyclonal antibody (Santa Cruz, Dallas, TX) and biotinylated anti-rabbit IgG (Sigma), both at 1:200 in Blocker A. For analyses of the mAbs, the arrayed slides were blocked with 3%



BSA in HBS with 5 mM CaCl<sub>2</sub> (Blocker B) and overlaid with the anti-H type 1 at 1:50 dilution; anti-H type 2, anti-A, anti-LNT, all at 1:100 dilution; anti-A type 2 at 1:200 dilution; anti-A type 1 and anti-A Lewis<sup>b</sup> culture supernatant undiluted. Binding was detected with biotinylated anti-mouse IgG (Sigma) at 1:200 dilution or anti-mouse IgM (Sigma,  $\mu$ -chain-specific) at 1:500 dilution. The biotinylated plant lectin UEA-I were analyzed using a single step overlay at 50  $\mu$ g/ml. For mAbs and lectins Blocker B was used as diluent. The final detection step was carried out by 30 min overlay with Alexa Fluor-647-labeled streptavidin (Molecular Probes, Eugene, OR, 1  $\mu$ g/ml in the corresponding blockers above).

Fluorescence images were recorded using ProScanArray microarray scanner (PerkinElmer LAS, Beaconsfield, UK) or GenePix 4300A (Molecular Devices, Sunnyvale, CA). For both scanners scanning resolution of 10  $\mu$ m/pixel was used and this resolution was adequate for the sizes of sample spots. Red laser (scan wavelength 633 nm) was used to record the protein binding signals. Images were analyzed using ScanArray Express software (PerkinElmer, Beaconsfield, UK) or GenePix® Pro 7 Microarray Analysis software (Molecular Devices, Workingham, UK), and the quantified gpr file was entered into an in-house microarray database using dedicated software (27) for data processing. No normalization method or statistical analysis was used for data processing, nor software or algorithms used for interpretation of processed data. The binding signals were dose related. Results are given for binding signals at 5 fmol per spot.

The above information is compliant with the MIRAGE (Minimum Information Required for A Glycomics Experiment) guidelines for reporting glycan microarray-based data (28) (SI Appendix).

**On-array Fucosidase Digestion**—For on-array glycosidase digestion, the array was hydrated with HEPES-buffered saline (HBS, 5 mM HEPES, pH 7.4, 150 mM NaCl, 5 mM CaCl<sub>2</sub>) for 2 min and blocked with 3% bovine serum albumin (BSA, Sigma) in HBS buffer for 1 h before addition of  $\alpha$ 1,2-fucosidase (New England Biolabs, Hitchin, UK, P0724L Lot 0111505) in 50 mM sodium citrate buffer, supplemented with 100 mM NaCl and 1  $\mu$ g/ml BSA (pH 6.0). The enzyme was evaluated at 8–20 unit/ $\mu$ l and incubation was at 37 °C for 12–44 h. The binding analyses were performed after treatment with the enzyme at 8–20 unit/ $\mu$ l. The array was then washed 3 times with HBS and once with water and dried at ambient.

## RESULTS

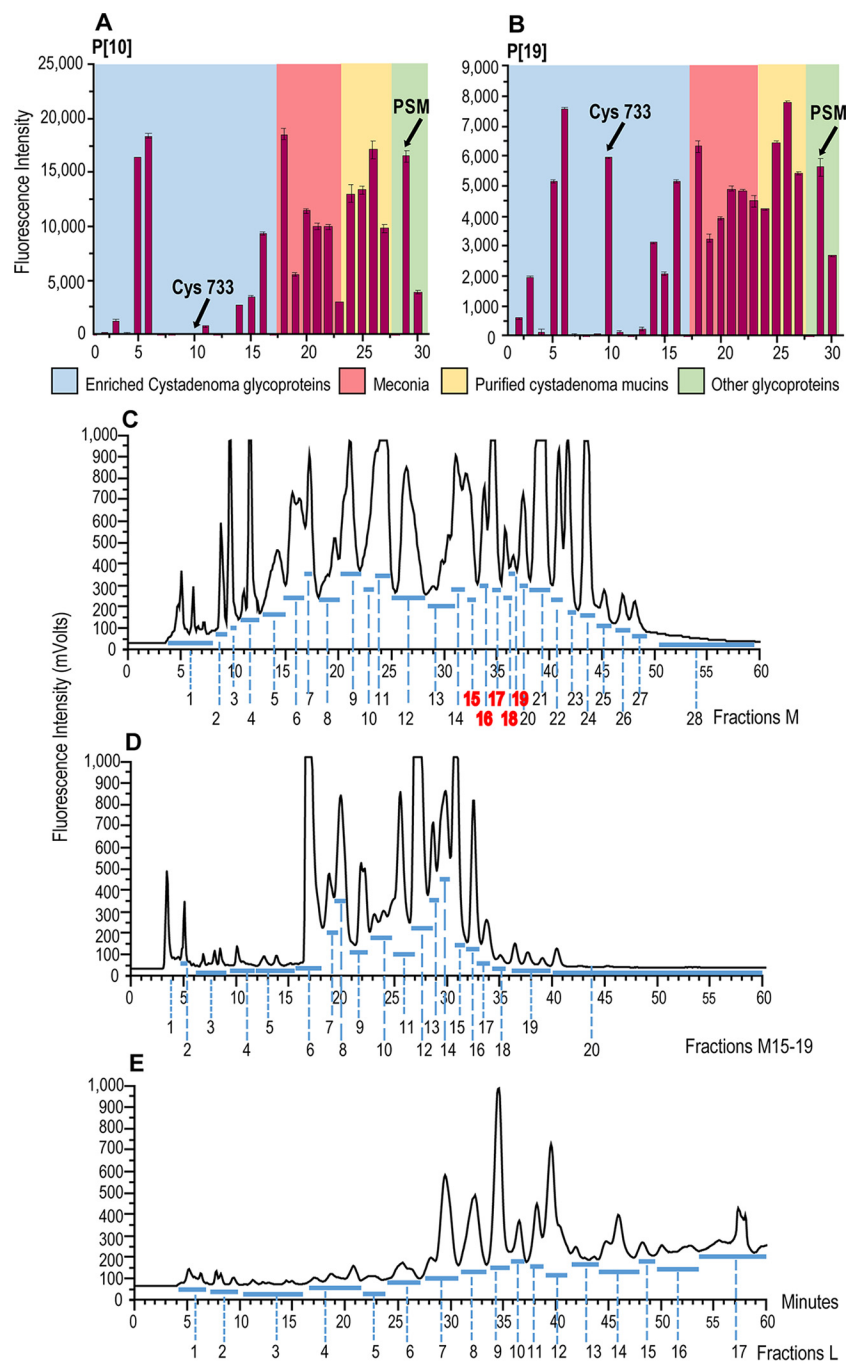
**Macromolecule Array for Selection of a Ligand-bearing Mucin Glycoprotein**—The work flow in the beam search approach (supplemental Fig. S1) began with the use of a microarray of 30 mucin-enriched glycoprotein preparations in order to select, as a study-case, a potent ligand-bearing mucin to reveal natural ligands therein for the VP8\* proteins of rotaviruses P[10] and P[19]. Among the mucins were those derived from human ovarian cystadenomas, meconia, and porcine stomach mucin (PSM) and bovine submaxillary mucin (BSM) (supplemental Table S1). There was binding by both VP8\* proteins to preparations of all five of the meconium-derived, eleven of the cystadenoma-derived mucins as well as to the PSM and BSM (Fig. 2A and 2B and supplemental Fig. S2). The binding patterns of the two viral proteins were similar overall, except that the P[19] VP8\* but not the P[10] VP8\* bound strongly to one additional cyst mucin, Cyst 733 (position 10, Fig. 2A and 2B, supplemental Table S1 and supplemental Fig. S2). Among the ligand-positive mucins, we se-

lected PSM for the O-glycome beam search for ligands for the P[10] and P[19] VP8\* proteins.

**Preparation of O-glycome and NGL Probes**—O-Glycans were reductively released from PSM, 1 g dry weight, and fractionated by gel filtration chromatography on a column of Bio-Gel P4 (supplemental Fig. S3). Thirteen pooled fractions (a to m) were collected. To have an overview of the O-glycan compositions the fractions were analyzed by MALDI-MS. Fractions a, b, c, l and m contained little or no detectable carbohydrate as indicated by MS-analyses and were not investigated further. In fractions d to k (total hexose content: 42 mg estimated to be equivalent to 84 mg sugar, considering that N-acetylhexosamines give no response in hexose assay (29)), O-glycan alditols with twenty-three molecular ions were detectable as major components, corresponding to sizes ranging from trisaccharide in fraction k to hexadecasaccharide in fraction d (supplemental Table S2). Undoubtedly there would be isomeric glycan species among these, as well as a multitude of additional minor components not detectable within the eight fractions under the MS conditions used.

Around 1 mol of each of the O-glycan alditol fractions d to k (estimate based on average molecular weights in the fractions; total sugar about 10 mg) was used for conversion to fluorescent NGLs (17) after mild periodate oxidation (18), and these were analyzed by high performance (HP) TLC (supplemental Fig. S4). The oxidative splitting of the core GalNAcol between C4 and C5 gives rise to two reactive aldehydes; one bearing the 6-linked glycan branch (if present) and the other 3-linked branch at GalNAcol (18) (supplemental Scheme S1). The monosaccharide compositions of the main NGL components detected in each of the fractions d to k are deduced from MALDI-MS analyses. At least twenty-three 3OX- molecular ions ranging from mono- to decasaccharides, and eleven 6OY-NGLs molecular ions ranging from mono- to hexasaccharides could be detected (supplemental Table S3).

The number of the 3OX-NGLs derived from each of the fractions (d to k) ranged from six to seventeen; the 6OY-NGLs were fewer, and ranged from four to eight. There were 3OX- and 6OY-NGL species with the same monosaccharide compositions in the eight fractions (unrelated to the sizes of the original O-glycan alditols). These represent chains 3-linked to core GalNAcol that may be accompanied by one or other of the 6-linked chains (supplemental Table S3). All of the fractions contained NGLs with the compositions H<sub>1</sub>-3OX, H<sub>1</sub>.N<sub>1</sub>-3OX and H<sub>1</sub>.N<sub>1</sub>-6OY; each of fractions f, g, i, j and k contained N<sub>1</sub>-6OY; and each of the fractions d, e, i, j and k contained dH<sub>1</sub>.H<sub>2</sub>.N<sub>1</sub>-3OX. These findings directed us to pool the NGLs of fractions d to k and thereby have similar molecules in one pot rather than in different fractions. The pool was fractionated and excess lipid reagent was removed by a small silica column (supplemental Fig. S5). NGL fractions with middle (M) and lower (L) migrations on TLC, which contain tri- to nonasaccharides and hexa- to undecasaccharides oligosaccha-



**FIG. 2. Steps in the preparation of the O-glycome beam search array.** (A) and (B) Microarray analyses of the VP8\* proteins of P[10] and P[19] enteroviruses at 25  $\mu\text{g/ml}$ , with a panel of 30 mucin-enriched glycoproteins (macromolecule array). C–E, Fractionation by HPLC of neutral O-glycome NGLs derived from PSM O-glycan alditols; the NGLs with middle migration patterns M, (described in supplemental Fig. S5) are in (C) and fractions are designated MHP1–MHP28; re-fractionation of the pooled fractions MHP15–19 is in (D) and the sub-fractions are designated  $M^{15-19}$ HP1– $M^{15-19}$ HP20; the NGLs with lower migration patterns, L, are in (E) and fractions are designated LHP1–LHP17. These NGL fractions constitute the primary array.

rides, respectively were combined and subjected to HPLC fractionation. The upper fraction which contained NGLs ranging from mono- to tri-saccharides was not pursued in present study as in the exploratory beam search it was found to be ligand-negative.

*Primary NGL Microarray to Identify Ligand-positive O-Glycome Fractions*—The HPLC-sub-fractions of the M and L fractions were designated MHP1 to MHP28 and LHP1 to LHP17, respectively (Fig. 2C and 2E). Based on exploratory experiments using 100 nmol from each of O-glycan fractions

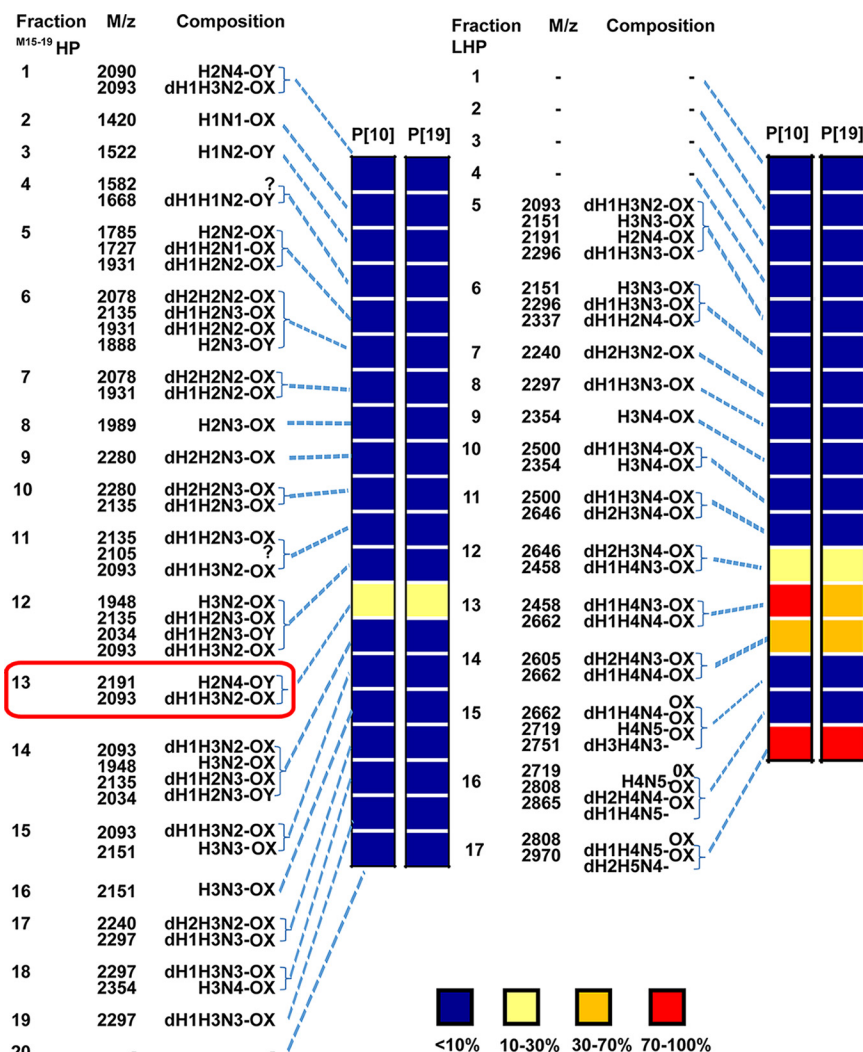


FIG. 3. Relative intensities of fluorescence in microarray analyses of the VP8\* proteins of rotaviruses P[10] and P[19] analyzed at 50  $\mu\text{g/ml}$  with neutral O-glycome NGL fractions  $M^{15-19}$ HP1 to  $M^{15-19}$ HP20 and LHP1 to LHP17. The molecular ions in MALDI-MS and deduced monosaccharide compositions of the fractions are also given. Inset, relative intensities are given with 100% for the highest signal for each protein. The fluorescence values for all of the arrayed neutral O-glycome NGL fractions that constitute the secondary array are in supplemental Table S4 and supplemental Fig. S7.

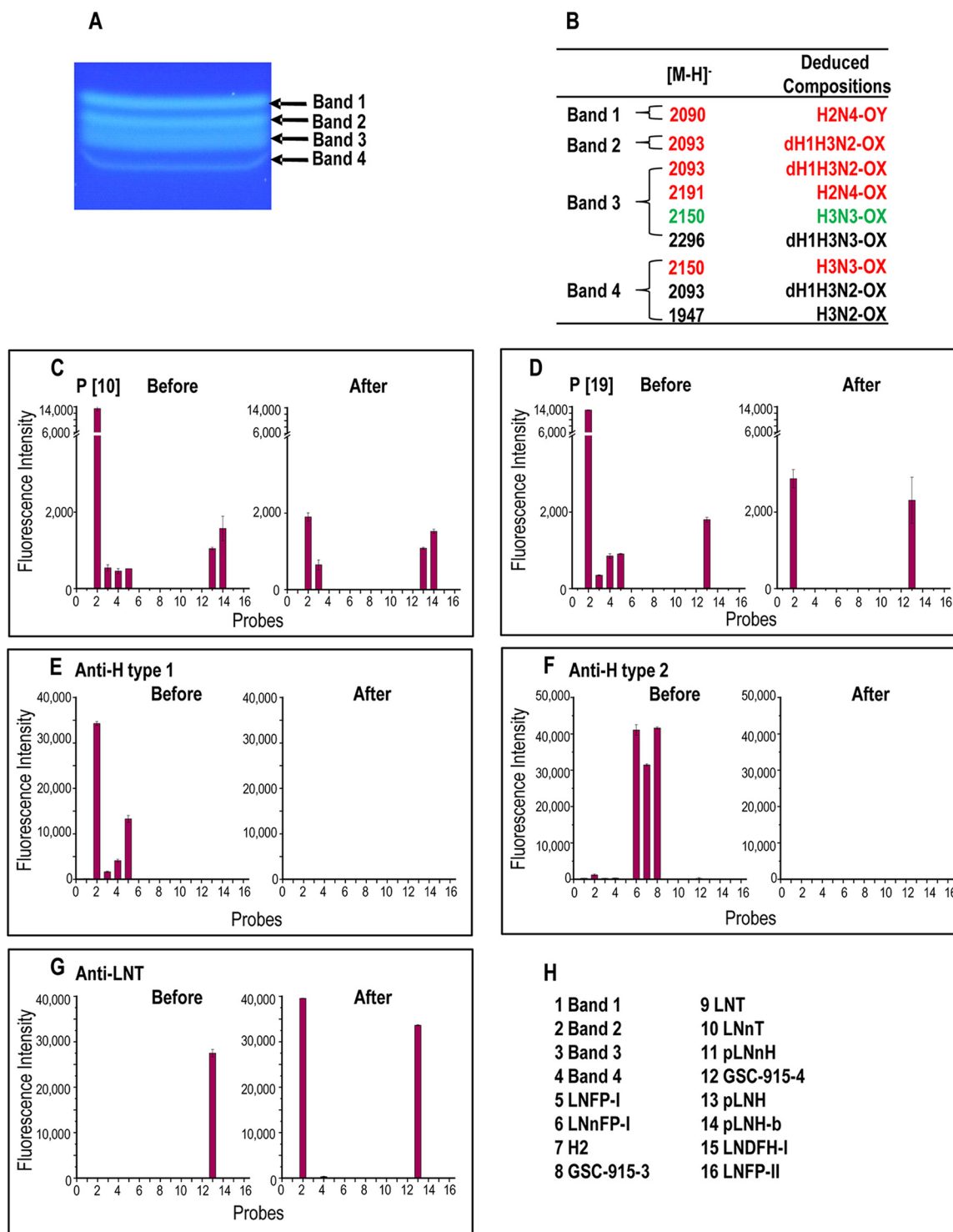
d to k, which showed that middle fraction 17 was ligand-positive for P[19]VP8\* (supplemental Fig. S6), we pooled fractions MHP15 to MHP19 and re-fractionated the pool using a new and better performing HPLC column. Twenty fractions were collected and designated  $M^{15-19}$ HP1 to  $M^{15-19}$ HP20 (Fig. 2D). The 60 fractions thus obtained (MHP1 to MHP14,  $M^{15-19}$ HP1 to  $M^{15-19}$ HP20, MHP20 to MHP28, and LHP1 to LHP17) were each analyzed by MALDI-MS, quantified and constituted the primary NGL microarray (supplemental Fig. S1 and supplemental Table S4). Five of the fractions were found to be ligand-positive for P[10] and P[19] VP8\* proteins. These were  $M^{15-19}$ HP13, LHP12, LHP13, LHP14 and LHP17 (Fig. 3 and supplemental Fig. S7).

To isolate and characterize in detail the glycan ligand(s) of the two VP8\* proteins we selected one of the more abundant of the four ligand-positive fractions,  $M^{15-19}$ HP13 (~ 8 nmol).

Two hexasaccharide-derived NGLs with compositions of  $dH_1H_3N_2-OX$  and  $H_2N_4-OY$  were detected by MS as the main components in this fraction (supplemental Table S4). It should be noted that  $dH_1H_3N_2-OX$  was also a component detected in the preliminary small scale study in the ligand positive middle fraction 17 (supplemental Fig. S6).

A quarter of the fraction  $M^{15-19}$ HP13 (~2 nmol) was sub-fractionated by HPTLC into four components (Fig. 4A) designated Bands-1 to -4 with approximate yields, 0.26, 0.26, 1.1, and 0.08 nmol, respectively. We extrapolate that the total yields of the four subfractions were 1.0, 1.0, 4.4 and 0.3 nmol, respectively.

*A Hexasaccharide Ligand Identified in the Secondary NGL Array*—A secondary NGL array was generated containing Bands 1–4, together with 12 reference glycans (Fig. 4H and supplemental Table S6). The predominant binding by the VP8\*



**FIG. 4. Fractionation by semi-preparative TLC, MALDI-MS analyses and immuno-sequencing of the components in ligand-positive fraction <sup>M15-19</sup>HP13.** *A*, High resolution semi-preparative TLC of <sup>M15-19</sup>HP13; Chromatography was upward; the solvent system was CHCl<sub>3</sub>/MeOH/Water 60:35:8 (by volume). *B*, Molecular ions and deduced monosaccharide compositions of the isolated Bands 1–4, panel (A); relative intensities of molecular ions are 80–100 red, 10–70 green, less than 10 black. *C* and *D*, Microarray analyses of the rotavirus P[10] and P[19] VP8\* proteins at 50 μg/ml. *E* to *G*, Microarray analyses of antibodies used as immuno-sequencing reagents: anti-H type 1, anti-H type 2 and anti-LNT. *H*, This secondary array (designated “Fucase Array”) included Bands 1–4 and the twelve reference compounds listed in panel *H*; their structures are given in supplemental Table S6. The analyses were performed before and after on-array treatment of the probes with α1–2 fucosidase. The results in *C* to *G*, are the means of fluorescence intensities of duplicate spots printed at 5 fmol/spot, with error bars representing half of the difference between the two values.



proteins was to Band-2 (Fig 4C and 4D). From MALDI-MS, Band-2 was assigned as consisting primarily of a single NGL dH<sub>1</sub>.H<sub>3</sub>.N<sub>2</sub>-3OX ([M-H]<sup>-</sup> at *m/z* 2093, Fig. 5A), and Band-1, H<sub>2</sub>.N<sub>4</sub>-6OY. Four components were detected in Band-3 and three components in Band-4. An NGL with dH<sub>1</sub>.H<sub>3</sub>.N<sub>2</sub>-3OX composition was among multiple components in Bands-3 and -4 (Fig. 4B). Binding analyses with the secondary NGL array showed strongest VP8\* binding to Band-2 (position 2 in Figs. 4C and 4D); there was only weak binding to Bands 3 and 4 (positions 3 and 4) and no binding to Band 1 (position 1).

Tandem MS of Band-2 was next performed (Fig. 5B). The sequential neutral loss of *m/z* 146, 162, 203, 162, 203, and 162 from *m/z* 2093 enabled the assignment of Band-2 as having a linear hexasaccharide sequence: dHex-Hex-Hex-NAC-Hex-HexNAC-Hex-3OX, derived from the 3-branch of the core O-GalNAcol.

*On-array Immuno-sequencing and MS of Ligands*—“Immuno-sequencing” of Band-2 was then carried out using monoclonal sequence-specific antibodies before and after on-array treatment with  $\alpha$ 1-2 fucosidase (Fig. 4). Band-2 was strongly bound by mAb 17-206 directed at the blood group H type 1 sequence: Fuc $\alpha$ 1-2Gal $\beta$ 1-3GlcNAc, but not by mAb BRIC231 directed at blood group H type 2 sequence: Fuc $\alpha$ 1-2Gal $\beta$ 1-4GlcNAc (Figs. 4E and 4F). Thus Band-2 was assigned as containing a blood group H type 1 trisaccharide-capping sequence, Fuc $\alpha$ 1-2Gal $\beta$ 1-3GlcNAc-Hex-HexNAC-Hex-3OX.

Microarray analyses after on-array treatment with  $\alpha$ 1-2 fucosidase showed loss of binding of Band-2 by both viral proteins as well as anti-H type 1 (Fig. 4C to 4E) and gain of binding by the anti-lacto-*N*-tetraose (LNT) (Fig. 4G). This enabled tentative assignment of a tandem “type 1-type 2” backbone sequence as below. Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal $\beta$ 1-4GlcNAc $\beta$ 1-Gal-3OX.

To our knowledge, the stringency of the anti-LNT for the type 1-type 2 backbone sequence above has not been established. To investigate this, we capitalized on the availability of the three reference compounds (Fig. 4H). These were pLNnH (position 11) with the “type 2-type 2” backbone sequence, pLNH (position 13) with the type 1-type 2 backbone sequence, and the novel compound pLNH-b (position 14) isolated in our laboratory (26). This isomer of pLNH has the “type 1-type 1” backbone sequence: Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc.

Among these three compounds, only pLNH was bound by the anti-LNT mAb (Fig. 4G), establishing the stringency of this antibody for the internal type 2 sequence in addition to the stated requirement for the outer type 1 (30). Thus the presence of type 1-type 2 backbone sequence in Band-2 could be unambiguously assigned.

However, the binding of one of the P[10] VP8\*, to pLNH-b sequence raised the possibility that Band 2 may contain a second component with the type 1-type-1 backbone sequence. An antibody specific for the pLNH-b sequence is not

available to our knowledge. This question could however be addressed by MALDI-CID-MS/MS analyses. For assignment of linkage types in the backbone of Band 2, NGLs of three isomeric hexasaccharides were analyzed as reference compounds to establish fragmentation patterns. These were: pLNnH with the type 2-type 2 backbone sequence, pLNH with the type 1-type 2”, and pLNH-b with the type 1-type 1 (Fig. 5C-5E). The three NGLs had identical molecular ions, [M-H]<sup>-</sup> at *m/z* 1720 and all three gave sequence ions at *m/z* 1558, 1355, 1193, 990, and 828, with sequential neutral loss of 162, 203, 162, 203, 162, and 162, respectively (in accordance with the linear sequence Gal-GlcNAc-Gal-GlcNAc-Gal-Glc). Only pLNH-b with “type 1-type 1” sequence gave a dehydrated ion (-h) at *m/z* 1175 (Fig. 5E). This dehydrated ion was more intense than its parent *m/z* 1193, and was totally lacking in pLNH and pLNnH with the type 2 internal linkage. This relatively intense dehydrated ion can serve as a marker for the internal type 1 linkage and be used as evidence for there being a pLNH-b type 1-type 1 linkage (26). The lack of the corresponding dehydrated ion at *m/z* 1402 arising from *m/z* 1420 (Fig. 5B) provided strong evidence for the lack of a second component with a type 1-type 1 backbone sequence in Band-2. Therefore, the backbone sequence of Band-2 was assigned as Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal $\beta$ 1-4GlcNAc $\beta$ 1-Gal $\beta$ 1-3OX.

Collectively the data enable the ligand-positive Band-2 to be assigned as a hexasaccharide with the type 1-terminating blood group H and an internal type 2 sequence, having been derived from a 3-linked branch at core GalNAcol: Fuc $\alpha$ 1-2Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal $\beta$ 1-4GlcNAc $\beta$ 1-Gal $\beta$ -3OX.

With both VP8\* proteins, the binding signal given by this hexasaccharide was substantially higher than that given by the pentasaccharide LNFP-I (at position 5, Fig 4C and 4D) implicating the involvement of the inner 4GlcNAc $\beta$ 1-Gal-sequence in the ligand for the two VP8\* proteins.

The four additional NGL fractions LHP12, LHP13, LHP14, and LHP17 containing larger glycans (8-11mers) that gave binding signals with the two VP8\* proteins (Fig. 3) were not investigated in detail; suffice it to say that there was immunochemical evidence for the presence of blood group H type 1 in all four fractions and blood group A type 1 in fractions LHP13, LHP14, and LHP17 (supplemental Table S5 and supplemental Fig. S7). Their monosaccharide compositions deduced by MALDI-MS (supplemental Table S4) are consistent with the immunochemical data. We comment further on these below, in the context of the binding data with a sequence-defined array.

*Validation of the Beam Search Approach with a Microarray of Sequence-defined Glycan Probes*—To validate the assignment of the ligand for the two VP8\* proteins identified within PSM neutral O-glycome using the beam search arrays, we performed microarray analyses using an array of sequence-defined glycan probes (Table I and supplemental Table S7). In complete accord with beam search array data, both P[10] and P[19] VP8\* proteins gave binding signals with the type 1 blood



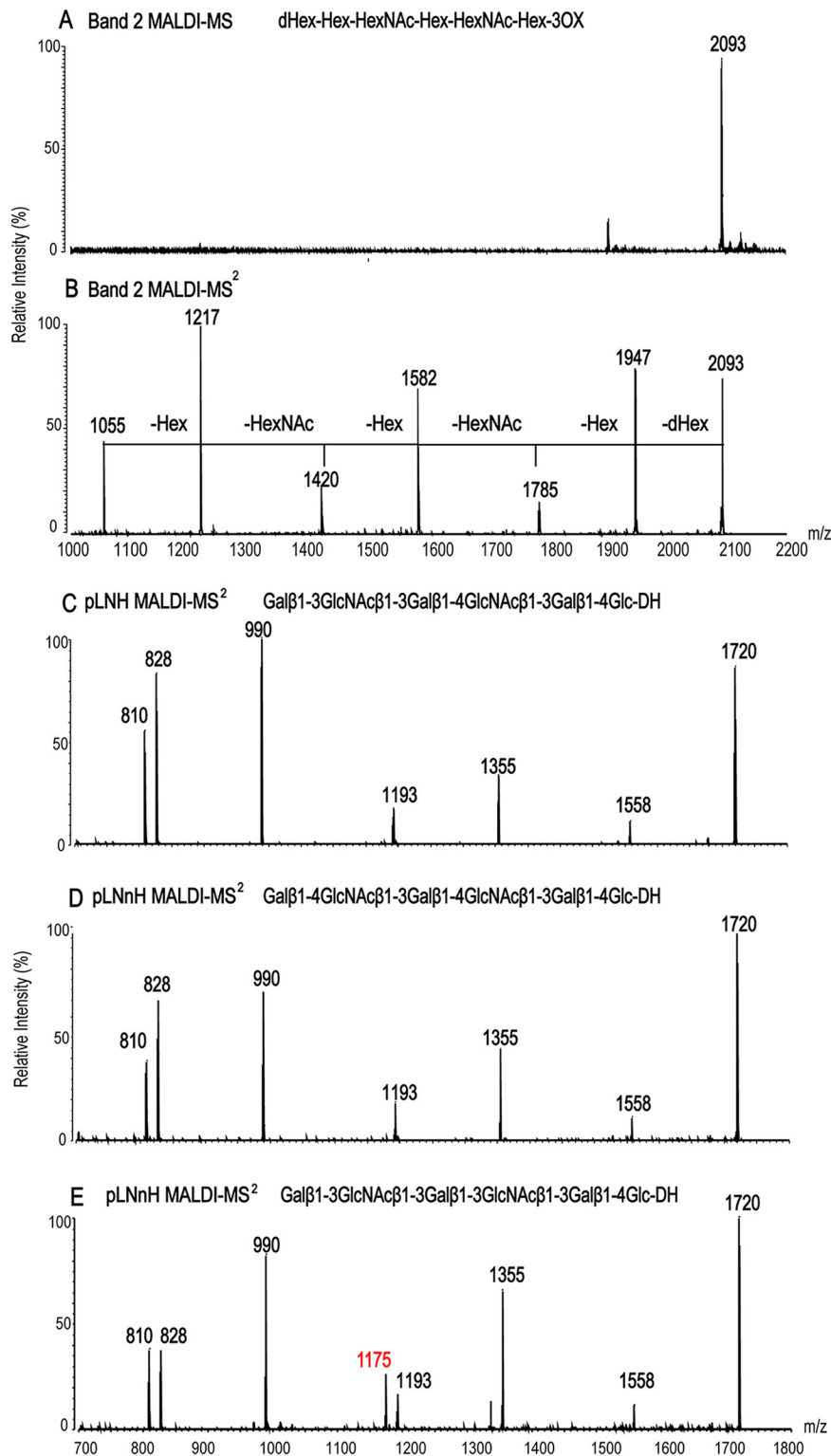


FIG. 5. MALDI-MS analyses of Band 2 and of the NGLs of reference compounds. A and B, MALDI-MS and CID-MS/MS of Band 2, respectively. C to E, MALDI-CID-MS/MS of NGLs of three isomeric hexasaccharides pLNH, pLNh and pLNh-b. used as reference compounds. The  $m/z$  ion in red in panel C is a dehydrated ion characteristic of the "type 1-type 1" backbone sequence in pLNH-b. The corresponding ions are lacking for pLNH and pLNh that have the "type 1-type 2" sequence, and also for Band 2.

TABLE I

Fluorescence intensities in microarray analyses of the VP8\* proteins of P[10] and P[19] (analyzed at 50 µg/ml) to probes selected for comparison in the F77/II array. Sequences with type 1 (Galβ1-3GlcNAc) and type 2 (Galβ1-4GlcNAc) terminating backbones, and terminal GlcNAc are shown in red, green, and blue font, respectively

Probe	Sequence <sup>a</sup>	P[10]	P[19]
		Fluorescence Intensity <sup>b</sup>	
LNFP-I	Fuα-2Galβ-3GlcNAcβ-3Galβ-4Glc-DH	668	1,628
LnNFPI	Fuα-2Galβ-4GlcNAcβ-3Galβ-4Glc-DH	- <sup>c</sup>	-
A-Hexa-T1	GalNAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH   Fuα-2	713	2,308
A-Hexa-T2	GalNAcα-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH   Fuα-2	-	-
A-Hepta	Fuα-4   GalNAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH   Fuα-2	-	3,636
B-Hexa-T1	Galα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH   Fuα-2	738	1,762
B-Hexa-T2	Galα-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH   Fuα-2	-	-
DFiLNO	Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6   Fuα-3   Galβ-4Glc-DH   Fuα-2Galβ-3GlcNAcβ-3	1,767	4,585
TFiLNO (1-2,2,3)	Fuα-2Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6   Fuα-3   Galβ-4Glc-DH   Fuα-2Galβ-3GlcNAcβ-3	9,945	18,663
pLNH	Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH	1,243	1,798
pLNnH	Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH	-	-
LNT	Galβ-3GlcNAcβ-3Galβ-4Glc-DH	-	-
LNnT	Galβ-4GlcNAcβ-3Galβ-4Glc-DH	-	-
Orsay-4	Galβ-3GlcNAcβ-3Gal-AO	55	-
Orsay-3	Galβ-3GlcNAcβ-6Gal-AO	-	1,384
Orsay-5	GlcNAcβ-6   Gal-AO   Galβ-3GlcNAcβ-3	68	13,668
O1.AO	GlcNAcβ-6   Gal-AO   GlcNAcβ-3	359	28,382

<sup>a</sup> DH-NGLs were prepared from reducing oligosaccharides by reductive amination with the amino lipid, 1,2-dihexadecyl-sn-glycero-3-phosphoethanolamine (DHPE); AO-NGLs were prepared from reducing oligosaccharides by oxime ligation with an aminoxy (AO) functionalized DHPE (40).

<sup>b</sup> Fluorescence intensities with probes printed at 5 fmol per spot.

<sup>c</sup> - Indicates fluorescence intensity less than 1.

group H sequence, LNFP-I (position 40). Two branched (bi-valent) forms, DFILNO and TFiLNO (1-2, 2, 3) included at positions 38 and 39, respectively, gave stronger binding signals than LNFP I. The binding observed additionally to the type 1 blood group A and B sequences in the sequence-defined array (Table I) is in accord with earlier data (25) reported while this work was underway. These data also support our conclusions that the type 1 blood group A antigen activities we detected in fractions LHP12, LHP13, LHP14, and LHP17 in the beam search array (Fig. 3 and supplemental Table S5) indeed arise from ligand-positive components in PSM O-glycome.

*Additional Specificities Revealed Using the Sequence-defined Screening Array*—The analyses using the sequence-defined glycan arrays revealed additional information on glycan sequences bound by the VP8\* proteins. (Table I and supplemental Table S7). These are glycan sequences not represented in PSM.

The first was the binding by both proteins to the non-fucosylated type 1-terminating (type 1-type 2) backbone sequence, pLNH (position 10), but not to the type 2-terminating (type 2-type2) analogue pLNnH (position 11). This observation is entirely consistent with the VP8\* protein binding observed after on-array fucosidase treatment of Band 2 (Fig. 4C and

4D). Binding of P[10] VP8\* to the pLNH-b, type 1-type-1 sequence was also shown (Fig. 4C).

Second, the ALe<sup>b</sup> sequence (position 46) was bound by the P[19] VP8\* but not that of the P[10]. Indeed among the mucins investigated, Cys 733, which was bound by the P[19] and not by the P[10] VP8\* protein strongly expresses ALe<sup>b</sup> (supplemental Fig. S2 and supplemental Table S1). We conclude that Le<sup>b</sup> sequence may be an important aspect of the difference between the two viruses. This is consistent with our earlier observations (25).

Third, there was pronounced binding by the P[19] but not the P[10] to short branched backbone sequences with a GlcNAc $\beta$ 1–6-linked to Gal (positions 13 and 14). Binding was considerably weaker to the unbranched GlcNAc $\beta$ 1–6-linked to Gal capped with Gal $\beta$ 1–3 (position 3).

Fourth was a finding of considerable interest unrelated to the blood group system, namely the binding of the P[19] but not the P[10] VP8\* to fungal cell wall type gluco-oligosaccharides (supplemental Table S8). These were of the curdlan and pustulan series ( $\beta$ 1–3-linked and  $\beta$ 1–6-linked sequences (positions 3–5 and 6–8, respectively) but not cellulose type:  $\beta$ 1–4-linked (position 9)).

#### DISCUSSION

We have adopted the term “beam search” from computer science where this terminology is used as a process or set of rules to be followed in calculations or problem-solving operations (31). The main features are ability to (a) find the answer or solution efficiently with limited memory requirements; (b) expand possible solutions within the beam width and search solutions or answers within the most promising data set and sub-data set; and (c) combine with other search methods for the ultimate answer (32). Our reasoning to designate the approach described here as beam search is that we aim to search as directly as possible for a ligand or ligands for a recognition protein using a population of natural O-glycans that are released from a mucin and may be available in limited amounts. To achieve this, we focus the search on the most promising fractions and sub-fractions. This has the advantage over the TLC plate binding approach (22) as the first line of analysis of the released O-glycans. It is of a substantially smaller scale and of a higher throughput as discussed below; and it allows binding analyses concomitant with MS monitoring of glycan content of the fractions during sequential chromatographies.

In the beam search approach, the macromolecule array analysis enables the selection of a ligand-positive, biologically relevant macromolecule (glycoprotein or polysaccharide) for ligand discovery. The primary NGL array is of fractions of released glycans and serves to identify ligand-positive glycan fractions while monitoring glycan compositions by MS. Iterative secondary, and if necessary subsequent arrays, are generated during successive purification steps. Microscale analysis methods are applied including MALDI-MS/MS and

immuno-sequencing before and after glycosidase treatments. In this prototype study, the selection of a mucin (PSM) from a macromolecule array, followed by just two glycan arrays enabled the identification and isolation of a ligand-positive hexasaccharide component recognized by the two viral proteins; these were an array of O-glycome NGL fractions resolved by HPLC followed by an array containing sub-fractions of a ligand-containing fraction resolved by HPTLC.

There are similarities in the aims of the beam search approach and that of the “shotgun” glycomics of Cummings and colleagues (33–36). Both are to characterize ligands among glycans from natural sources, using sequence analyses by MS/MS and applying sequence-specific lectins and antibodies. There are differences, however, in the design and application of the two approaches. The shotgun glycomics is to present the total glycome from a natural source so that it can be screened to identify physiologically or biologically relevant glycans. The dissection of the glycome is by multidimensional HPLC fractionation aiming to separate individual glycans “as homogeneous as possible” in the first place for printing. Biological questions and applications often come at a later stage. In the beam search array approach, on the other hand, a ligand-bearing macromolecule for a specific recognition system is identified, from which to release, pinpoint and isolate the natural glycan ligand(s); the process involves a ligand-guided iterative fractionation and arraying/probing process focusing on the ligand-positive components which are also the most biologically relevant in the system and context investigated. Hence the generation of a primary and a secondary glycan array from that glycome to cone down on a component expressing the ligand.

The ligand-positive component Band-2, ~1 nmol of NGL that we isolated from PSM (containing ~1  $\mu$ g carbohydrate), amounted to ~0.2% of the 600 nmol NGLs obtained from 10 mg of O-glycan alditols. This was sufficient for unambiguous structural characterization and incorporation into microarrays for up to 8000 binding analyses in our current microarray format. With the chromatogram-binding system, which requires a minimum of 100 pmol of NGLs per lane, only 10 binding analyses can be performed. Thus, the beam search approach is orders of magnitude more sensitive as well as being of higher throughput. The microarray data in the small-scale experiments, using as little as ~1 mg of O-glycan alditols, showed that ligand-containing fractions can be readily detected by this approach with associated MALDI-MS monitoring. The ligand-positive hexasaccharide with the blood group H type 1 sequence (rather than H type 2) has not been described in PSM to our knowledge (37). With optimization of the O-glycan derivatization step there is scope for a substantial increase of yields.

The beam search approach calls for sequential fractionation and pooling steps (supplemental Fig. S1). The O-glycans released are first size-fractionated prior to oxidative cleavage and lipid conjugation, so that glycans of differing sizes can be separately converted to NGLs, using our established micro-

scale conjugation protocols (17, 18), considering average molecular weights of the glycans in the fractions. This also facilitates the MS survey of the variously sized glycans released. After visualizing the resulting NGLs by high performance TLC they are pooled ahead of various chromatographies. The rationale for pooling the NGL population at this stage is to have in one-pot the 3OX- and 6OY-NGLs of similar sizes that arise following the oxidative cleavage of the core GalNAcol of the original branched O-glycans of differing sizes. There follows block fractionation of the NGL population using silica gel cartridge before separations by HPLC, and robotic arraying to generate the primary arrays for analyses with carbohydrate-binding proteins of interest. Additional fractionation methods, are used at this stage as appropriate. While considering the advantages of the simplification of glycan profile following the core GalNAcol splitting after the mild periodate oxidation, mention is required that this is at the expense of much of the natural O-glycan core which is largely sacrificed. Work is under way to develop further and broaden the beam search technology to encompass core-intact O-glycans released by a recently described method (38). This will serve to identify core-specific recognition which may be missed following the core-split. The core-intact, branched O-glycans will however have their native heterogeneity and require rigorous fractionations.

The new results from the beam search approach applied to PSM, namely the assignment of the blood group H type 1 sequence as well as the non-fucosylated type 1 backbone sequence as ligands for the two VP8\* proteins of the P[10] and P[19] viruses are fully validated by microarray analyses with the sequence-defined glycans. Moreover, the comparative data showed preferential binding of the VP8\* proteins to the O-glycan component isolated (Band-2) that has a pentasaccharide backbone relative to LNFP I with a tetrasaccharide backbone. Also shown was the ability of the two VP8\* proteins to accommodate the blood group A and B determinants (GalNAc $\alpha$ 1–3 and Gal $\alpha$ 1–3, respectively). These findings now account for the lack of a correlation between secretor status and the binding of the two viral proteins to salivary mucins in which the expression of the blood group antigens is controlled by the secretor gene-associated blood group H enzyme (25) (supplemental Table S1).

The additional ligand sequences revealed for the two VP8\* proteins by the analyses with the sequence-defined screening arrays are of considerable interest. The three sequences identified that were bound by the P[19] VP8\* protein but not by that of the P[10] could potentially come to play in biological contexts other than the pig gastrointestinal tract: The first is the ALe<sup>b</sup> sequence which is prominent in one of the human epithelial mucins (cyst 733) (supplemental Fig. S2) and would be predicted to occur in the gastrointestinal intestinal epithelial glycoproteins of humans who are blood A Lewis-positive secretors<sup>4</sup>. The second sequence preferentially bound by the P[19] VP8\* protein is GlcNAc $\beta$ 1–6Gal; this is a candidate terminal motif on poly-N-acetylglucosamine chains of O- and

N-glycans as well as glycolipids unrelated to blood group determinants and secretor status. It will be interesting to investigate the distribution of this motif in tissues as well as possible cross-reactions with the mucin core 2, 4 and 6 sequences that were shown earlier to be bound by the P[19] VP8\* protein (25). The third P[19]-associated specificity is non-mammalian; it is toward the fungal polysaccharide-type  $\beta$ 1–3 and  $\beta$ 1–6 linked glucose sequences. This finding raises the intriguing possibility of interactions of the virus with fungi in the intestinal microbiome where involvement in regulatory processes termed “transkingdom” interactions of which there is an increasing awareness (39). These findings show the important complementarity of the increasing repertoire of sequence-defined glycan arrays with the arrays generated *de novo* from natural sources.

Our immuno-sequencing data highlight the scope for expanding the antibody-tools concomitantly with analyses using sequence-defined glycan arrays to provide reference compounds not only to determine specificities of binding, but also to reveal unsuspected recognition structures that are represented in different physiological contexts.

In conclusion, the beam search approach described here paves the way to O-glycome recognition studies in a wide range of basic and medical settings to give new insights into glycan recognition structures in natural microenvironments. A macromolecule array followed by a primary NGL array of the derived, fractionated O-glycome NGLs can indeed serve as the starting point for beam search of thousands of diverse binding systems that recognize a given mucin macromolecule.

*Acknowledgments*—The sequence-defined glycan microarrays contain many saccharides provided by collaborators whom we thank as well as members of the Glycosciences Laboratory for their collaboration in the establishment of the neoglycolipid-based microarray system. We thank Henrik Clausen for providing murine anti-A type 1 and anti-A Lewis<sup>b</sup> antibodies and Atsushi Hara and Dr. Akihiro Imamura for synthesis of glycan GSC-915; Mark S. Stoll for advice on microarray data handling and Colin Herbert for technical assistance.

#### DATA AVAILABILITY

The authors declare that the data supporting the findings of this study are compliant with MIRAGE (MIRAGE Glycan Microarray Guidelines - Beilstein-Institut) and are available within the paper and its supplementary information files and Appendix.

\* This work was supported by Wellcome Trust (WT 108430/Z/15/Z). ZL was supported in part by the China Scholarship Council.

§ This article contains supplemental material.

§§ To whom correspondence should be addressed: Glycosciences Laboratory, Department of Medicine, Imperial College London, W12 0NN, UK. E-mail: t.feizi@imperial.ac.uk or w.chai@imperial.ac.uk.

‡‡ ZL and CG contributed equally to this work.

#### REFERENCES

1. Andrianifahanana, M., Moniaux, N., and Batra, S. K. (2006) Regulation of mucin expression: Mechanistic aspects and implications for cancer and inflammatory diseases. *BBA-Rev. Cancer* **1765**, 189–222



2. Tran, D. T., and Ten Hagen, K. G. (2013) Mucin-type O-Glycosylation during Development. *J. Biol. Chem.* **288**, 6921–6929
3. Arike, L., and Hansson, G. C. (2016) The Densely O-Glycosylated MUC2 Mucin Protects the Intestine and Provides Food for the Commensal Bacteria. *J. Mol. Biol.* **428**, 3221–3229
4. van Kooyk, Y., and Rabinovich, G. A. (2008) Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat. Immunol.* **9**, 593–601
5. Robbe, C., Capon, C., Coddeville, B., and Michalski, J. C. (2004) Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem. J.* **384**, 307–316
6. Pinho, S. S., and Reis, C. A. (2015) Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555
7. Cummings, R. D., and McEver, R. P. (2009) C-type Lectins. In: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E., eds. *Essentials of Glycobiology 2<sup>nd</sup> Ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
8. Rossez, Y., Maes, E., Darroman, T. L., Gosset, P., Ecobichon, C., Curt, M. J. C., Boneca, I. G., Michalski, J. C., and Robbe-Masselot, C. (2012) Almost all human gastric mucin O-glycans harbor blood group A, B or H antigens and are potential binding sites for *Helicobacter pylori*. *Glycobiology* **22**, 1193–1206
9. Revoredo, L., Wang, S. J., Bennett, E. P., Clausen, H., Moremen, K. W., Jarvis, D. L., Ten Hagen, K. G., Tabak, L. A., and Gerken, T. A. (2016) Mucin-type O-glycosylation is controlled by short- and long-range glycopeptide substrate recognition that varies among members of the polypeptide GalNAc transferase family. *Glycobiology* **26**, 360–376
10. Kiessling, L. L., and Splain, R. A. (2010) Chemical approaches to glycobiology. *Annu. Rev. Biochem.* **79**, 619–653
11. Krasnova, L., and Wong, C. H. (2016) Understanding the Chemistry and Biology of Glycosylation with Glycan Synthesis. *Annu. Rev. Biochem.* **85**, 599–630
12. Watkins, W. M. (1980) Biochemistry and Genetics of the ABO, Lewis, and P blood group systems. *Adv. Hum. Genet.* **10**, 1–136, 379–385
13. Kabat, E. A. (1982) Philip Levine Award Lecture. Contributions of quantitative immunochemistry to knowledge of blood group A, B, H, Le, I and I antigens. *Am. J. Clin. Pathol.* **78**, 281–292
14. Tang, P. W., Gool, H. C., Hardy, M., Lee, Y. C., and Feizi, T. (1985) Novel approach to the study of the antigenicities and receptor functions of carbohydrate chains of glycoproteins. *Biochem. Biophys. Res. Commun.* **132**, 474–480
15. Yuen, C. T., Lawson, A. M., Chai, W. G., Larkin, M., Stoll, M. S., Stuart, A. C., Sullivan, F. X., Ahern, T. J., and Feizi, T. (1992) Novel sulfated ligands for the cell-adhesion molecule E-selectin revealed by the neoglycolipid technology among O-linked oligosaccharides on an ovarian cystadenoma glycoprotein. *Biochemistry* **31**, 9126–9131
16. Yuen, C. T., Chai, W. G., Loveless, R. W., Lawson, A. M., Margolis, R. U., and Feizi, T. (1997) Brain contains HNK-1 immunoreactive O-glycans of the sulfoglucuronyl lactosamine series that terminate in 2-linked or 2,6-linked hexose (mannose). *J. Biol. Chem.* **272**, 8924–8931
17. Stoll, M. S., Feizi, T., Loveless, R. W., Chai, W., Lawson, A. M., and Yuen, C. T. (2000) Fluorescent neoglycolipids. Improved probes for oligosaccharide ligand discovery. *Eur. J. Biochem.* **267**, 1795–1804
18. Chai, W., Stoll, M. S., Galustian, C., Lawson, A. M., and Feizi, T. (2003) Neoglycolipid technology: deciphering information content of glycome. *Methods Enzymol.* **362**, 160–195
19. Fukui, S., Feizi, T., Galustian, C., Lawson, A. M., and Chai, W. (2002) Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions. *Nat. Biotechnol.* **20**, 1011–1017
20. Liu, Y., Childs, R. A., Palma, A. S., Campanero-Rhodes, M. A., Stoll, M. S., Chai, W., and Feizi, T. (2012) Neoglycolipid-based oligosaccharide microarray system: preparation of NGLs and their noncovalent immobilization on nitrocellulose-coated glass slides for microarray analyses. *Methods Mol. Biol.* **808**, 117–136
21. Palma, A. S., Feizi, T., Childs, R. A., Chai, W. G., and Liu, Y. (2014) The neoglycolipid (NGL)-based oligosaccharide microarray system poised to decipher the meta-glycome. *Curr. Opin. in Chem. Biol.* **18**, 87–94
22. Gao, C., Liu, Y., Zhang, H. T., Zhang, Y. B., Fukuda, M. N., Palma, A. S., Kozak, R. P., Childs, R. A., Nonaka, M., Li, Z., Siegel, D. L., Hanfland, P., Peehl, D. M., Chai, W. G., Greene, M. I., and Ten, F. Z. (2014) Carbohydrate sequence of the prostate cancer-associated antigen F77 assigned by a mucin O-glycome designer array. *J. Biol. Chem.* **289**, 16462–16477
23. Huang, P., Xia, M., Tan, M., Zhong, W., Wei, C., Wang, L., Morrow, A., and Jiang, X. (2012) Spike protein VP8\* of human rotavirus recognizes histo-blood group antigens in a type-specific manner. *J. Virol.* **86**, 4833–4843
24. Liu, Y., Huang, P., Tan, M., Liu, Y., Biesiada, J., Meller, J., Castello, A. A., Jiang, B., and Jiang, X. (2012) Rotavirus VP8\*: phylogeny, host range, and interaction with histo-blood group antigens. *J. Virol.* **86**, 9899–9910
25. Liu, Y., Ramelot, T. A., Huang, P., Liu, Y., Li, Z., Feizi, T., Zhong, W., Wu, F. T., Tan, M., Kennedy, M. A., and Jiang, X. (2016) Glycan Specificity of P(19) Rotavirus and Comparison with Those of Related P Genotypes. *J. Virol.* **90**, 9983–9996
26. Gao, C., Zhang, Y., Liu, Y., Feizi, T., and Chai, W. (2015) Negative-ion electrospray tandem mass spectrometry and microarray analyses of developmentally regulated antigens based on type 1 and type 2 backbone sequences. *Anal. Chem.* **87**, 11871–11878
27. Stoll, M. S., and Feizi, T. (2009) Software Tools for Storing, Processing and Displaying Carbohydrate Microarray Data After. In: Kettner, C., ed. *Beilstein Symposium on Glyco-Bioinformatics*, pp. 123–140, Beilstein Institute for the Advancement of Chemical Sciences Potsdam, Germany
28. Liu, Y., McBride, R., Stoll, M., Palma, A. S., Silva, L., Agravat, S., Aoki-Kinoshita, K. F., Campbell, M. P., Costello, C. E., Dell, A., Haslam, S. M., Karlsson, N. G., Khoo, K. H., Kolarich, D., Novotny, M. V., Packer, N. H., Ranzinger, R., Rapp, E., Rudd, P. M., Struwe, W. B., Tiemeyer, M., Wells, L., York, W. S., Zaia, J., Kettner, C., Paulson, J. C., Feizi, T., and Smith, D. F. (2016) The minimum information required for a glycomics experiment (MIRAGE) project: improving the standards for reporting glycan microarray-based data. *Glycobiology* **27**, 280–284
29. Anderson, K., Li, S. C., and Li, Y. T. (2000) Diphenylamine-aniline-phosphoric acid reagent, a versatile spray reagent for revealing glycoconjugates on thin-layer chromatography plates. *Anal. Biochem.* **287**, 337–339
30. Bry, L., Falk, P. G., and Gordon, J. L. (1996) Genetic engineering of carbohydrate biosynthetic pathways in transgenic mice demonstrates cell cycle-associated regulation of glycoconjugate production in small intestinal epithelial cells. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1161–1166
31. Ow, P. S., and Morton, T. E. (1988) Filtered beam search in scheduling. *Int. J. Prod. Res.* **26**, 35–62
32. Furcy, D., and Koenig, S. (2005) Limited discrepancy beam search. *19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, 125–131
33. Smith, D. F., and Cummings, R. D. (2014) Investigating virus-glycan interactions using glycan microarrays. *Curr. Opin. Virol.* **7**, 79–87
34. Yu, Y., Lasanajak, Y., Song, X. Z., Hu, L. Y., Ramani, S., Mickum, M. L., Ashline, D. J., Prasad, B. V. V., Estes, M. K., Reinhold, V. N., Cummings, R. D., and Smith, D. F. (2014) Human milk contains novel glycans that are potential decoy receptors for neonatal rotaviruses. *Mol. Cell Proteomics* **13**, 2944–2960
35. Song, X., Lasanajak, Y., Xia, B., Heimburg-Molinaro, J., Rhea, J. M., Ju, H., Zhao, C., Molinaro, R. J., Cummings, R. D., and Smith, D. F. (2011) Shotgun glycomics: a microarray strategy for functional glycomics. *Nat. Methods* **8**, 85–90
36. Byrd-Leotis, L., Liu, R., Bradley, K. C., Lasanajak, Y., Cummings, S. F., Song, X., Heimburg-Molinaro, J., Galloway, S. E., Culhane, M. R., Smith, D. F., Steinhauer, D. A., and Cummings, R. D. (2014) Shotgun glycomics of pig lung identifies natural endogenous receptors for influenza viruses. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2241–E2250
37. Cheng, P. F., Snovidia, S., Ho, M. Y., Cheng, C. W., Wu, A. M., and Khoo, K. H. (2013) Increasing the depth of mass spectrometry-based glycomic coverage by additional dimensions of sulfoglycomics and target analysis of permethylated glycans. *Anal. Bioanal. Chem.* **405**, 6683–6695
38. Song, X., Ju, H., Lasanajak, Y., Kudelka, M. R., Smith, D. F., and Cummings, R. D. (2016) Oxidative release of natural glycans for functional glycomics. *Nat. Methods* **13**, 528–534
39. Pfeiffer, J. K. (2016) Host response: Microbiota prime antiviral response. *Nat. Microbiol.* **1**, 15029
40. Liu, Y., Chai, W., Childs, R. A., and Feizi, T. (2006) Preparation of neoglycolipids with ring-closed cores via chemoselective oxime-ligation for microarray analysis of carbohydrate-protein interactions. *Methods Enzymol.* **415**, 326–340