Glycomics and Proteomics Analyses of Mouse Uterine Luminal Fluid Revealed a Predominance of Lewis Y and X Epitopes on Specific Protein Carriers*

Chu-Wei Kuo‡§, Chin-Mei Chen§, Ying-Chu Lee§, Sin-Tak Chu§¶, and Kay-Hooi Khoo§**

Sperm motility and maturation are known to be affected by a host of factors encountered en route in both male and female genital tracts prior to fertilization. Using a concerted proteomics and glycomics approach with advanced mass spectrometry-based glycan sequencing capability, we show in this work that 24p3, an abundant mouse uterine luminal fluid (ULF) glycoprotein also called lipocalin 2 (Lcn2), is highly fucosylated in the context of carrying multiple Lewis X and Y epitopes on complex type N-glycans at its single glycosylation site. The predominance of Lewis X/Y along with Neu5Ac2–6 sialylation was found to be a salient feature of the ULF glycome, and several other protein carriers were additionally identified including the highly abundant lacto-transferrin, which is N-glycosylated at two sites, both with a similar range of highly fucosylated N-glycans. A comparative glycomics analysis of the male genital tract fluids revealed that there is a gradient of glycomic complexity from the cauda to caput regions of the epididymis, varying from high mannose to sialylated complex type N-glycans but mostly devoid of fucosylation. The seminal vesicle fluid glycome, on the other hand, carries equally abundant multimeric Lewis X structures but is distinctively lacking in additional fucosylation of the terminal galactose to give the Lewis Y epitope typifying the glycome of female ULF. One-dimensional shotgun proteomics analysis identified over 40 proteins in the latter, many of which are reported for the first time, and a majority are notably involved in immune defense and antigen processing. Further sperm binding and motility assays suggest that the Lewis X/Y epitopes do contribute to the sperm motility-enhancing activity of 24p3, whereas lactotransferrin is largely inactive in this context despite being similarly glycosylated. These findings underline the importance of glycoproteomics in delineating both the specific glycan structures and their carriers in assigning glycobiological functions. Molecular & Cellular Proteomics 8:325–342, 2009.

The advent of functional genomics and proteomics has facilitated systematic investigations into the key events from spermatogenesis to fertilization in reproductive glycobiology using the mouse model (1). Although the general importance of glycosylation is well recognized (2, 3), specific structural details of the implicated glycotopes are surprisingly lacking. The best studied glycotopes are those on the zona pellucida glycoproteins (4–6) thought to mediate sperm binding to oocytes. Virtually nothing is known about the potential roles mediated by glycosylation as the maturing spermatozoa moves from storage in epididymis through ejaculation during copulation into the uterine tract, encountering en route a myriad of secretory glycoproteins within the bathing fluids. Adsorption or interactions with various glycotopes presented by these secretory glycoproteins may have profound effect in modulating the characteristics of sperm glycocalyx (7, 8) and, through that, its fertilizing potential. The observation that glycotopes are often involved in eliciting antisperm immune responses (9) and the many reported incidences of their dynamic modulation in association with sperm fertility underscore the importance of precise structural and functional knowledge in efforts directed toward developing them as targets for immuncontraceptive vaccines (10).

In human, glycodelin is a major uterine fluid and seminal plasma glycoprotein with gender-specific glycosylation and contraceptive properties (11, 12). It belongs to the lipocalin superfamily characterized by shared three-dimensional structure and the ability to bind small hydrophobic molecules (13). The progesterone-regulated glycodelin-A of uterine origin potently and dose-dependently inhibits human sperm-egg binding through interaction with sperm, whereas differently glycosylated glycodelin-S from seminal plasma has no such effect (14). The latter is unusually fucose-rich, carrying a high abun-
Glycoproteomics of Mouse Uterine Luminal Fluid

Fucosylation is widely implicated in mouse reproductive physiology from epididymal sperm maturation to blastocyst implantation (18, 19). In particular, expression of α2-fucosylated glycans on the uterine epithelium is known to be dynamically regulated in concert with the estrous cycle during which the α2-fucosyltransferase mRNA level is high after estrogenic stimulation and ovulation but becomes negligible after implantation (20). Such hormonal regulation of glycosylation is expected to extend to the synthesized and secreted components. A gradient of fucosyltransferase activity has also been demonstrated to exist from the caput to cauda epididymis that parallels the apparent acquisition by the maturing spermatozoa of surface α2-fucosylated glycans and the ability to fertilize during epididymal passage (19, 21). Optimal exposure and density of these glycoproteins are dynamically remodeled upon further interactions with uterine secretory components as the sperm attains timely capacitation in the female reproductive tract. Nonetheless the respective functions of specific fucosylated epitopes remain unsolved. Deficiency in the expression of α2-fucosylated glycans on either the epididymal or uterine epithelia alone in mice with targeted deletions of the Fut1 or Fut2 α2-fucosyltransferase, respectively, did not seem to impair fertility (22). On the other hand, mice with a null mutation in the FX locus, which encodes an enzyme involved in the de novo pathway for GDP-fucose synthesis, were infertile when reared in the absence of fucose (23), thus supporting an essential role for fucosylation in reproduction. Likewise about half of the components not known previously to be associated with the uterine luminal fluid and show that Leα and Leβ are the major fucosylated epitopes of its constitutive glycoproteins including 24p3 and lactotransferrin. The same Leβ epitope, which is also the predominant glycotope on human glycodelin-S but not glycodelin-A, is further demonstrated here to be involved in sperm/24p3 interactions. Interestingly only Leα but not Leβ was found on the male seminal vesicle fluid glycoproteins, whereas similar glycomics profiling of the epididymal fluids revealed that neither of the fucosylated glycoproteins is presented to the maturing sperm in these sections of the male reproductive tracts.

Experimental Procedures

Preparation of 24p3, Uterine Luminal Fluid (ULF), Seminal Vesicle Fluid (SVF), Epididymal Fluid (EF), and Spermatozoa—ULF was collected from 3-week-old female mice pretreated with diethylstilbestrol, and 24p3 glycoprotein was purified thereof as described previously (16). The seminal vesicles of normal adult mice (8–12 weeks old) were carefully dissected to free them from the adjacent coagulating glands, and the secretions were squeezed directly into ice-cold 5% acetic acid. After stirring for 30 min at 4 °C, the fraction was centrifuged at 3500 × g for 20 min at 4 °C, and the supernatant was collected as SFV. The epididymides from normal adult mice (8–12 weeks old) were isolated and divided into caput, corpus, and cauda. Each segment was minced in 1.5 ml of PBS buffer and gently shaken for 10 min to permit dispersal of the luminal content. After the tissue pieces were allowed to settle, the upper fraction containing sperm and luminal fluid was collected and centrifuged at 3500 × g for 20 min at 4 °C. The supernatant was then collected as EF. The total protein content of each fluid sample was determined by Bradford assay using the Bio-Rad Protein Assay kit prior to further analyses.

For the preparation of spermatozoa, the caudal segments of the epididymides were minced and incubated for 10 min at 37 °C in a modified HEPES medium (HM) described previously (17) for sperm preparation that contained 120 mM NaCl, 2 mM KCl, 1.2 mM MgSO4, 7H2O, 0.36 mM NaH2PO4, 25 mM NaHCO3, 10 mM HEPES, 5.6 mM glucose, 1.1 mM sodium pyruvate, 100 IU/ml penicillin, and 100 μg/ml streptomycin. The pH was adjusted to a level of 7.3–7.4 by aeration with humidified air/CO2 (19:1) in an incubator at 37 °C for 4 h prior to use. Spermatozoa that were extruded from the distal portion of this tissue were filtered through a nylon gauze and collected by centrifugation at 60 × g for 10 min at room temperature with three repeated washings in the same medium.

SDS-PAGE and Lectin Blot—For lectin blot analysis, the SDS-PAGE gel was first rinsed in transfer buffer containing 192 mM glycine, 20% (v/v) methanol, and 25 mM Tris-HCl, pH 8.3, for 10 min and then transferred onto PVDF membrane (Millipore, Billerica, MA) using a Hoefer Semi-Dry Transfer Unit (Amersham Biosciences) at 120 mA for 1 h. Nonspecific binding was blocked with 3% (w/v) BSA in 10 mM Tris-HCl, pH 7.5, 100 mM NaCl, and 0.1% (v/v) Tween 20 (TBST) for 1 h at room temperature before applying the biotinylated lectins for

1 The abbreviations used are: Leα, Lewis X; Leβ, Lewis Y; AAL, A. aurantia lectin; DES, diethylstilbestrol; EF, epididymal fluid; LacNac, N-acetyllactosamine; SVF, seminal vesicle fluid; ULF, uterine luminal fluid; Lcn2, lipocalin 2; Fuc, fucose; Hex, hexose; HexNAc, N-acetyllactosamine; HM, HEPES medium; NCBlnr, National Center for Biotechnology Information non-redundant; PID, precursor ion discovery; GC, gas chromatography; Leα, Lewis A; Leβ, Lewis B; 1D, one-dimensional; NeuAc (Neu5Ac), N-acetyllactosamine acid; Neu5Ac, 5-N-acetyllactosamine acid; Neu5Gc, 5-N-glycolylnuraminic acid; LPS, lipopolysaccharide.
Glycoproteomics of Mouse Uterine Luminal Fluid

1 h at room temperature. The *Aleuria aurantia* lectin and *Lotus tetragonolobus* lectin (Vector Laboratories, Burlingame, CA) were used to dilute 1:800 in TBST with 1% (w/v) BSA. After washing with TBST, the membranes were incubated with horseradish peroxidase-conjugated streptavidin (Vector Laboratories) diluted 1:10,000 in TBST with 1% (w/v) BSA for another 1 h at room temperature and subsequently developed using the ECL detection system (Amersham Biosciences).

**Gel-based MS/MS Analysis of Peptides for Protein Identification**—The ULF protein bands detected on SDS-PAGE by SYPRO Ruby (Molecular Probes, Eugene, OR) staining were manually excised from the gel, reduced, alkylated, trypsin-digested, and subjected to nano-LC-nano-ESI-MS/MS analysis on a Q-TOF Ultima™ API (Micromass, Manchester, UK) instrument for protein identification as described previously (25). Raw data files were processed by ProteinLynx 2.2 (smoothing method: Savitzky-Golay, six channels, twice; center method: four channels, centroid top 80%) and converted into pik files for searching against an NCBI database (June 29, 2007; 5,162,317 sequences) using an in-house Mascot Daemon 2.2 server. The criteria for database searching were set as follows: species specified as *Mus musculus* (146,831 sequences), trypsin digestion, variable modification as carbamidomethyl (Cys) and oxidation (Met), allowance of up to one missed cleavage, and mass accuracy of 50 ppm on the parent ion and 0.25 Da on the fragment ions. The proteins identified as significant hits by Mascot (p < 0.05, protein score >32) were further filtered based on a protein score >50 contributed by at least two peptides identified at a score >30 with all individual peptide ion scores >15 and reproducibly identified in repeated runs.

**MALDI MS/MS Analysis of De-N-glycosylated Peptides**—Selected protein bands including those corresponding to 24p3 and lactotransferrin were additionally subjected to manual MALDI MS and CID MS/MS analyses on a dedicated Q-TOF Ultima MALDI instrument (Micromass) before and after N-glycosidase F (Roche Applied Science) digestion of the extracted tryptic peptides. Peptides subjected to overnight de-N-glycosylation treatment at 37 °C in 50 mM ammonium bicarbonate, pH 8.5, were desalted and separated from the released N-glycans by using a C₁₈ ZipTip (Millipore, Bedford, MA). Bound peptides were eluted off the ZipTip with 75% acetonitrile and 1% formic acid after repeated washing with 0.1% formic acid, dried down, redissolved in 50% acetonitrile and 0.1% formic acid, and mixed 1:1 with matrix solution (5 mg/ml α-cyano-4-hydrocinnamic acid and 2% ammonium citrate in the same solvent) for spotting onto the MALDI sample stage. Glu-fibrinopeptide B was used as near point lock mass. MALDI MS/MS de novo sequencing data were manually interpreted against the known protein sequence in addition to searching against the database with an additional criterion of Asn to Asp conversion to verify the implicated N-glycosylation site.

**In-solution Digestion and Shotgun LC-MS/MS Analysis of Peptides and Glycopeptides**—Total ULF proteins were reduced with 55 mM tris(2-carboxyethyl)phosphine (AttoChem, Madison, WI) at an enzyme to substrate ratio of 1:100 (w/w) at 37 °C overnight all in the same 25 mM ammonium bicarbonate buffer, pH 8.5. The enzyme digestion was stopped by adding 0.1% formic acid and further diluted with acetonitrile as needed for direct shotgun analysis.

**Shotgun Proteomics Identifications**—Nano-LC-nano-ESI-MS/MS analysis was performed on a Paradigm MS4 nanoflow system (Michrom BioResources, Auburn, CA) connected to an LTQ-Orbitrap XL hybrid mass spectrometer (Thermo Electron, Bremen, Germany) equipped with a PicoView nanospray interface (New Objective, Woburn, MA). Peptide mixtures were loaded onto a 75-μm × 250-mm fused silica capillary column packed in house with C₁₈ resin (5 μm, Nucleosil 120-5 C₁₈, Macherey-Nagel, GmbH & Co. KG) and were separated using a segmented gradient in 100 min from 5 to 65% solvent B (95% acetonitrile with 0.1% formic acid) at a flow rate of 200 nl/min. Solvent A was 0.1% formic acid in water. The mass spectrometer was operated in the data-dependent mode. Briefly survey full-scan MS spectra were acquired in the orbitrap (m/z 350–1600) with the resolution set to 60,000 at m/z 400 and automatic gain control target at 10⁵. The 10 most intense ions were sequentially isolated for CID MS/MS fragmentation and detection in the linear ion trap (automatic gain control target at 7000) with previously selected ions dynamically excluded for 90 s. Ions with single and unrecognized charge states were also excluded. All the measurements in the orbitrap were performed with the lock mass option for internal calibration. The MS and MS/MS raw data were processed using the Extract_Msn module in Bioworks 3.3.1 (molecular weight range, 600–6000; grouping tolerance set to 0; precursor charge set to auto; minimum scans per group; and intermediate scans set to 1) and searched against an in-house generated NCBI database comprising mouse ORFs in their forward and reversed orientations (February 18, 2008; 475,610 sequences) using an in house Mascot Daemon 2.2 server. Search criteria used were trypsin digestion, variable modifications set as carbamidomethyl (Cys) and oxidation (Met), allowance of up to one missed cleavage, and mass accuracy of 10 ppm on the parent ion and 0.6 Da on the fragment ions. All significant protein hits from Mascot (p < 0.05, protein score >32) thus obtained contained no false positive hit from the reverse database. This resulting data set was then subjected to further statistical analysis using the Trans-Proteomic Pipeline with PeptideProphet and ProteinProphet probability cutoffs set at 0.05 and 0.9, respectively, which filtered the reported protein hit data set to an estimated false positive error rate of 0.7%, at an 87.1% sensitivity. Gene ontology categories for the identified proteins were assigned by using the DAVID (Database for Annotation, Visualization, and Integrated Discovery) bioinformatics resources and used to classify the proteins according to their biological processes and molecular functions.

**Shotgun Analysis of Glycopeptides in Precursor Ion Discovery Mode**—Independent nano-LC-nano-ESI-MS/MS analysis of the solution-based total tryptic digests targeting specifically the glycopeptides was performed on the Micromass Q-TOF Ultima API mass spectrometer using the conditions as described above for gel-based analysis. To facilitate identification of glycopeptides, automated MS/MS data-dependent acquisition was operated under the precursor ion discovery (PID) mode (26). In brief, two successive LC-MS survey scans with the collision cell set at alternating low (7 eV) and high (30 eV) collision energies were used to trigger MS/MS acquisition on the five most intense parent ions observed at the low energy survey scans when glycan-specific oxonium ions fragments, m/z 204.084 for HexNAc⁺ and m/z 366.139 for HexHexNAc⁺, were detected at the corresponding high energy scans. MS/MS acquisition on false positives was limited to a single scan if the monitored oxonium ions were not afforded to devote more analysis time on true positives. MS/MS data were manually verified, and the corresponding MS survey scans encompassing the eluting time for the entire range of glycoforms were delineated and summed based on total and selected ion chromatograms.

**Lectin Enrichment and Shotgun Proteomics Identification of De-N-glycosylated Peptides**—Alternatively the tryptic peptides from a total of 150 μg of ULF proteins were desalted using a C₁₈ Sep-Pak cartridge and subjected to lectin enrichment using the *A. aurantia* lectinagarose beads (Vector Laboratories) equilibrated in buffer containing 10 mM Tris-Cl, pH 7.5, and 150 mM NaCl. Peptide sample dissolved in the equilibration buffer was incubated overnight at 4 °C with the lectin beads. Adsorbed glycopeptides were eluted off the agarose beads with the same buffer containing 100 mM L-Fuc after repeated washing and desalted using a C₁₈ Sep-Pak cartridge (Waters, Milford,
MA). After further overnight N-glycosidase F treatment in 50 mm ammonium bicarbonate, pH 8.5, at 37 °C, peptide samples were again desalted and separated from the released N-glycans by C18 Sep-Pak cartridge (Waters) and subjected to shotgun proteomics identification by nano-LC-nano-ESI-MS/MS analysis on the Micromass Q-TOF Ultima API mass spectrometer as described above for gel-based analysis but with the added criterion of Asn to Asp conversion as one of the variable modifications for database searching. MS/MS spectra were further examined and verified manually for reliability of glycosylation site assignment.

MALDI MS and MS/MS Glycomics Analyses—For glycomics analyses of purified 24p3, ULF, EF, and SVF, samples were redissolved in 50 mm ammonium bicarbonate, pH 8.4; first reduced with dithiothreitol at 37 °C for 4 h; and then alkylated with iodoacetamide at room temperature for 4 h in the dark followed by removal of excess reagents by passing through a Sep-Pak C8 cartridge (Waters). Samples were then digested with trypsin (Roche Applied Science) overnight at 37 °C followed by N-glycosidase F after brief boiling and cooling down. Released N-glycans were separated from peptides by passing through C18 Sep-Pak (Waters). For additional desialylation, the N-glycans were digested with 300 000 units of neuraminidase from Arthrobacter ureafaciens (Roche Applied Science) in 25 µl of 50 mm sodium acetate buffer, pH 5.0, for 24 h at 37 °C and desalted by passing through a mixed bed ion exchange column packed with 1 ml each of Dowex (50W-X8, 50–100 mesh, protonated form, Bio-Rad) and AG 3-X4 (AG 3-X4, 100–200 mesh, free base form, Bio-Rad) resins.

All released N-glycans, including those from gel-based derived tryptic peptides described above, were permethylated using a modified NaOH/DMSO method (27) prior to MS analysis. GC-MS linkage analysis on partially methylated alditol acetates prepared thereof was performed as described previously (28). For MALDI-TOF MS glycan profiling, the permethyl derivatives in acetonitrile were mixed 1:1 with 2,5-dihydroxybenzoic acid matrix (10 mg/ml in acetonitrile), spotted on the target plate, air-dried, and recrystallized on plate with acetonitrile. Data acquisition was performed manually on a benchtop MiLDI LR system (Micromass) or the 4700 Proteomics Analyzer (Applied Biosystems, Framingham, MA) operated in the reflectron mode. MALDI MS/MS sequencing of the permethylated glycans was performed on both the Q-TOF Ultima MALDI and 4700 Proteomics Analyzer exactly as described before (29).

Protein Labeling and Flow Cytometric Analysis—FITC-labeled protein was prepared by incubating purified protein (5 mg/ml) with FITC (5 mg/ml) in 100 mm sodium bicarbonate buffer, pH 9.0, for a period of 20 h at 4 °C. Unconjugated FITC was removed by a Bio-Spin C6 column (Bio-Rad). For flow cytometric analysis, the freshly prepared spermatozoa (10⁵ cells/ml) were incubated in HM at 37 °C for 30 min with 0.2 nmol of FITC-labeled protein (FITC-24p3 or FITC-lactotransferrin) and 2 nmol of competitive proteins (purified 24p3 or lactotransferrin from ULF). Flow cytometry was performed on a Coulter EPICS XL flow cytometer (Beckman-Coulter, Miami, FL). Fluorescence was initiated by excitation at a wavelength of 488 nm and measured via a 525-nm filter. Photomultiplier tube voltages and gains were set to maximize the dynamic range of the signal. The fluorescence intensity of the FITC and protein complex was quantified for 10,000 individual cells.

Measurement of Sperm Motility—The motility and movement characteristics of spermatozoa were assessed following incubation of the sperm in the modified HM medium at a concentration adjusted to 10⁶ cells/ml with and without addition of 24p3, lactotransferrin, and other small oligosaccharides (Leα, Leβ, lacto-N-neotetraose, blood group H type II, and 6-sialyl-N-acetyllactosamine from Dextra Laboratories Ltd., Berkshire, UK) at various concentrations. For the mouse model, the characteristics of the trajectories of single spermatozoan have been studied, and various swimming patterns have been described (10, 11). The parameters relating to the spermatozoa motility were analyzed by computer-assisted sperm assay with a sperm motility analyzer (IVOS (Integrated Video Optical System) version 10, Hamilton-Thorne Research, Beverly, MA). A 10-µl sample was placed in a 10-µm-deep Markler chamber at 37 °C. The analyzer was set as follows: negative phase-contrast optics and recording at 60 frames/s, minimum contrast set at 40, minimum cell size set to four pixels, low size gate set to 0.2, high size gate set to 1.5, low intensity gate set to 0.5, high intensity gate set to 1.5, non-motile head size set to 29, non-motile head intensity set to 76, medium average path velocity adjusted to 50 µm/s, low average path velocity adjusted to 7.0 µm/s, slow motile cells selected, and threshold strictness set to a level greater than 80%. Fifteen fields were assessed for each sample.

RESULTS

Glycomics Profiling of ULF Proteome and 24p3—To obtain a first impression of the overall N-glycosylation pattern of the ULF glycoproteome, several batches of the released total N-glycan pools were permethylated for glycomics mapping by MALDI MS. Although slight variations in the relative intensity of individual peaks were registered, the profiles were consistently dominated by two major series of sialylated molecular ions (Fig. 1A). The first corresponds to the ubiquitous high mannose structures, Man₃₋₄(GlcNAc)₂, (m/z 1580, 1784, 1988, 2192, and 2396), whereas the second could be assigned as multifucosylated bi-, tri-, and tetraantennary complex type N-glycans with up to two Fuc residues per antenna in addition to core fucosylation. Those underfucosylated components were also visibly accompanied by Neu5Ac-monomosialylated counterparts; albeit they were relatively minor in the overall glycomic profile. In comparison, the monosialylated components were more prominent among the N-glycans released from purified 24p3, which represents a major ULF constituent. After permethylation, the substitution of two Fuc residues by a Neu5Ac would give a mass increment of 13 Da. It is obvious from the glycomic profile (Fig. 1B) that most of the major peaks occurred in pairs of 13 amu apart, thus indicative of multifucosylation and monosialylation. These pairings disappeared upon neuraminidase treatment, which removed the single Neu5Ac sialylation and thereby collapsed the heterogeneity, revealing a full extent of fucosylation on multiantennary N-glycans with up to five LacNAc units (Fig. 1C).

MS/MS Sequencing Defined Leα and Leβ as Major Terminal Epitopes—To determine the fine structural details of the assigned molecular compositions, the major signals detected were subjected to complementary modes of MS/MS analyses as reported previously (29). By low energy CID on a MALDI-Q-TOF instrument, the distribution of Fuc on terminal epitopes versus core structures could be easily defined by the abundant sialodiated B and Y ions arising from cleavages at the GlcNAc. Thus, a Neu5Ac-sialylated Hex-HexNAc, a non-fucosylated Hex-HexNAc, and that with one and two Fuc residues were identified by the sialodiated B ions at m/z 847, 486, 660, and 834, respectively, and further corroborated by the Y
FIG. 1. MALDI MS profiles of permethylated N-glycans from total proteins of the mouse ULF (A) and purified 24p3 thereof before (B) and after (C) neuraminidase treatment. All labeled [M + Na]+ molecular ion signals could be tentatively assigned as annotated, and the structures of major components were further verified by MS/MS analyses. Insets in B and C represent magnified m/z regions to better show the minor components at high mass range. ○, Hex (light gray, Gal; dark gray, Man); ■, HexNAc (GlcNAc); ◆, Neu5Ac; ▼, Fuc. High mannose type N-glycans were annotated simply as Man₅₋₉, representing Man₅₋₉(GlcNAc)₂.
ions arising from consecutive losses of these non-reducing terminal epitopes from the parent ions (data not shown). The assignment of these fragment ions is usually straightforward and allows one to rapidly establish the overall structures. However, in general, linkage-specific cleavages are not detected to enable assignment of the critical terminal epitopes.

In contrast, high energy CID MS/MS spectra afforded by MALDI-TOF/TOF is often complicated by an array of fragment ions arising from cross-ring cleavages and concerted elimination around the ring structures, which nevertheless can be systematically assigned to provide a definitive mapping of the linkage positions. In particular, the relatively high abundance D ions are sufficiently diagnostic to define a terminal Lewis structure as exemplified by the representative MS/MS data shown in Fig. 2. In the case of a trifucosylated biantennary structure (Fig. 2A), the D ions formed at the GlcNAc for Lex and Ley were detected at m/z 472 and 646, respectively, whereas the corresponding ion for either Lea or Leb was not found at m/z 442 (30). Other ions as assigned and schematically illustrated unambiguously defined the two isomeric structures. Likewise for the sialylated structure the D ion for Ley was detected at m/z 646 (Fig. 2B). In this case, the D ion
formed at the core β-Man (m/z 1228 accompanied by the 3,5A ion at m/z 1126) further showed that the Le^v epitope was preferentially carried on the 6-arm. In addition, the Neu5Ac was defined as 6-linked to Gal by the D ion at m/z 558, which would not have formed from the alternative 3-linked structure (29). Further confirmation was provided by conventional GC-MS linkage analysis of the total permethylated glycan pool showing that 6-linked Gal was present at relatively high abundance (data not shown) with no appreciable amount of 3-linked Gal. The linkage analysis data also supported the Ley structure by the presence of 2-linked Gal.

From the MALDI MS mapping and MS/MS sequencing data supplemented by GC-MS linkage analysis, it could thus be concluded that the major complex type structures detected are indeed core fucosylated structures substituted by terminal Le^a and Le^v without any evidence for the presence of 1 chain epitopes Le^a and Le^v. Sialylation, where it occurs, is mostly α2–6 monosialylation by Neu5Ac on a non-fucosylated LacNAc unit and therefore does not constitute a terminal sialyl Lewis epitope. This overall terminal glycosylation pattern of the ULF glycome and that of 24p3 are very similar. The latter nevertheless carries a relatively higher amount of the monosialylated structures and very little of the high mannose structures. Biantennary complex type structures are consistently the predominant component series, but structures carrying up to five LacNAc units and eight Fuc residues were detected within the mass range and sensitivity limit. For each of the bi-, tri-, and tetraantennary series, heterogeneity resides mainly on the degree of fucosylation, which collectively accounts for a high Fuc content of 24p3 (16) and the ULF glycoproteome in general.

The Proteomic Content of ULF and Common Glycosylation Pattern—Because the N-glycosylation pattern of 24p3 was found to be very similar to the N-glycomic profile of ULF, it would suggest that either the N-glycans of 24p3 are the major constituents of the overall ULF glycome or that most other major ULF glycoproteins similarly carry the same repertoire of N-glycan structures with a prevalence of terminal Le^v/Le^v epitopes. To distinguish between the two possibilities, a multifaceted glycoproteomics approach in concert with direct proteomics identifications was undertaken. The proteomic content of ULF was first “visualized” by 1D SDS-PAGE in conjunction with LC-MS/MS analyses of the in-gel tryptic digests of each of the major protein bands (Fig. 3A, Table I, and supplemental Data 1A). This allowed subsequent correlation of the lectin blot and glycosylation analysis data with the identified major constituent proteins. A more comprehensive identification of the ULF proteome was additionally performed by a one-dimensional shotgun LC-MS/MS analysis of the total tryptic digest, which resulted in a total of 44 proteins identified with high confidence (ProteinProphet value >0.9 with estimated false positive rate 0.7%; supplemental Data 1B and 2) and that may be classified according to their molecular functions or involved biological processes based on gene ontology annotation (Fig. 3B).

When cross-referenced against the 1D gel-based data, it is apparent that a characteristic of the ULF proteome is the predominance of complement C3 and its various degraded forms as it could be identified from almost every excised band with high scores. Not surprisingly, this and most other proteins identified are one way or the other involved in immune defense and/or antigen processing (see “Discussion”), and a majority of these proteins are not previously known to be associated with ULF. Furthermore it can be inferred from the 1D gel band intensities that 24p3 (lipocalin 2) and lactotransferrin are indeed among the two most abundant ULF proteins. For a rapid assessment of the individual glycosylation patterns, N-glycans were released in gel from these and other
Glycoproteomics of Mouse Uterine Luminal Fluid

TABLE I
Identification of major mouse ULF proteins from each of the 1D gel bands by nano-LC-MS/MS analysis

<table>
<thead>
<tr>
<th>Gel band</th>
<th>Protein name</th>
<th>Accession number</th>
<th>Molecular weight (theoretical)</th>
<th>Mascot score</th>
<th>Matched peptides</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aminopeptidase N</td>
<td>gi</td>
<td>31077182</td>
<td>109,582</td>
<td>482</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>Complement factor H precursor</td>
<td>gi</td>
<td>116132</td>
<td>138,992</td>
<td>364</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>Complement C3 precursor</td>
<td>gi</td>
<td>1352102</td>
<td>186,365</td>
<td>292</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>α 3A chain of laminin-5</td>
<td>gi</td>
<td>1922889</td>
<td>190,271</td>
<td>115</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>Complement C3 precursor</td>
<td>gi</td>
<td>1352102</td>
<td>186,365</td>
<td>443</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>FcgGBP</td>
<td>gi</td>
<td>21410127</td>
<td>127,072</td>
<td>186</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>Ceruloplasmin isofrom b</td>
<td>gi</td>
<td>110347564</td>
<td>121,074</td>
<td>170</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>Complement C3 precursor</td>
<td>gi</td>
<td>1352102</td>
<td>186,365</td>
<td>398</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>Amiloride-binding protein 1 (amine oxidase, copper-containing)</td>
<td>gi</td>
<td>22095013</td>
<td>85,396</td>
<td>275</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>Complement factor B</td>
<td>gi</td>
<td>6996919</td>
<td>84,951</td>
<td>175</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>Alkaline phosphatase</td>
<td>gi</td>
<td>22138740</td>
<td>57,452</td>
<td>130</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>BC048546 protein</td>
<td>gi</td>
<td>21707647</td>
<td>45,301</td>
<td>117</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>Lactotransferrin precursor</td>
<td>gi</td>
<td>21431833</td>
<td>77,816</td>
<td>1148</td>
<td>24</td>
</tr>
<tr>
<td>14</td>
<td>Complement C3 precursor</td>
<td>gi</td>
<td>1352102</td>
<td>186,365</td>
<td>1255</td>
<td>23</td>
</tr>
<tr>
<td>15</td>
<td>Lactotransferrin precursor</td>
<td>gi</td>
<td>21431833</td>
<td>77,816</td>
<td>453</td>
<td>8</td>
</tr>
<tr>
<td>16</td>
<td>Serum albumin precursor</td>
<td>gi</td>
<td>5915682</td>
<td>68,648</td>
<td>242</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>Complement C3 precursor</td>
<td>gi</td>
<td>1352102</td>
<td>186,365</td>
<td>271</td>
<td>7</td>
</tr>
<tr>
<td>18</td>
<td>Chloride channel calcium-activated 3</td>
<td>gi</td>
<td>8567363</td>
<td>100,009</td>
<td>250</td>
<td>6</td>
</tr>
<tr>
<td>19</td>
<td>Complement factor B</td>
<td>gi</td>
<td>6996919</td>
<td>84,951</td>
<td>237</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>CD14 antigen</td>
<td>gi</td>
<td>6753332</td>
<td>39,179</td>
<td>155</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>Complement C3 precursor</td>
<td>gi</td>
<td>1352102</td>
<td>186,365</td>
<td>841</td>
<td>18</td>
</tr>
<tr>
<td>22</td>
<td>Chitinase 3-like protein 1</td>
<td>gi</td>
<td>33468847</td>
<td>42,952</td>
<td>192</td>
<td>4</td>
</tr>
<tr>
<td>23</td>
<td>Cathepsin B preproprotein</td>
<td>gi</td>
<td>6681079</td>
<td>37,256</td>
<td>96</td>
<td>3</td>
</tr>
<tr>
<td>24</td>
<td>Complement factor B</td>
<td>gi</td>
<td>6996919</td>
<td>84,951</td>
<td>96</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>Complement C3 precursor</td>
<td>gi</td>
<td>1352102</td>
<td>186,365</td>
<td>379</td>
<td>10</td>
</tr>
<tr>
<td>26</td>
<td>Lipocalin 2</td>
<td>gi</td>
<td>34328049</td>
<td>22,860</td>
<td>380</td>
<td>5</td>
</tr>
<tr>
<td>27</td>
<td>Complement C3 precursor</td>
<td>gi</td>
<td>1352102</td>
<td>186,365</td>
<td>177</td>
<td>3</td>
</tr>
</tbody>
</table>

**Notes:**
- Gel band numbers refer to those shown in Fig. 3A. Only the most abundant proteins confidently identified from each of the protein bands are listed. Related proteins identified by the same unique peptide matches are reported as a single entry represented by the most descriptive protein name.
- Number of unique peptide matches with an ion score threshold of 15. Full details of the individual peptide matches and scores are provided in supplemental Data 1A.

major bands, permethylated, and profiled by MALDI MS. It was found that the resulting spectra were fairly similar with respect to the major glycan signals but with slight variations in their relative intensities (data not shown). More relevant to the focus of this work is the conclusion that the multiply fucosylated N-glycan pattern in the forms of terminal Leα/Leβ is common to all protein bands investigated and not unique to 24p3.

Identification of Specific ULF Proteins Carrying the Multifucosylated N-Glycans—Because each major protein gel band in general contains more than a single glycoprotein (Table I), it is not possible to attribute the identified multifucosylated N-glycan profiles to a particular protein carrier amid the presence of other glycoforms. To this end, identification of specific glycopeptides carrying highly fucosylated N-glycans would be necessary. An initial attempt at direct shotgun LC-MS/MS analysis of the total ULF tryptic peptide pool using the PID mode for data-dependent acquisition succeeded only in unambiguous identification of the two glycopeptides corresponding to the two predicted N-glycosylation sites of lactotransferrin (Fig. 4). Two glycopeptide clusters eluting at two different time points that carried the characteristic glycosyl residual mass differences between signals within each cluster were clearly identified from manual examination of the MS survey spectra (Fig. 4, A and B). Automated MS/MS acquisition was triggered on several of the signals. As shown in Fig. 4, C and D, for the ESI-MS/MS spectra of the glycopeptides representing each of the inferred sites, confirmation of a true glycopeptide signal was first afforded by the characteristic presence of the highly abundant singly charged glycan-specific oxonium ions at m/z 204 (HexNAc−), 274 and 292 (NeuAc−), 366 (Hex,HexNAc−), 512 (Fuc,Hex,HexNAc−), 657
(NeuAc\textsubscript{1}Hex\textsubscript{1}HexNAc\textsuperscript{+}), and 658 (Fuc\textsubscript{2}Hex\textsubscript{1}HexNAc\textsuperscript{+}). In addition, sequential loss of glycosyl residues from the parent ions could be observed and manually assigned, leading to identification of putative signals corresponding to the peptide core and peptide core retaining a single HexNAc. All signals within the same glycopeptide cluster gave the common doubly charged peptide core + HexNAc signals at either \(m/z\) 915.92 or 973.48 when the respective MS/MS spectra were of sufficiently good quality. Notably the peptide cores manually deduced from the MS/MS spectra matched exactly those assigned from independent MALDI MS/MS analyses of the de-\(N\)-glycosylated tryptic peptides derived from the identified lactotransferrin gel band. Two peptides at \(m/z\) 1628 and 1743 were detected in the MALDI MS spectrum only after de-\(N\)-glycosylation, and manual de novo sequencing clearly established the requisite conversion of the implicated Asn at the consensus site to Asp (data not shown).

Unambiguous identification of other glycopeptides based on direct LC-MS/MS analysis was not possible because, in general, peptide sequence-informative \(b\) and \(y\) ions were hardly detectable, and delineation of the MS/MS spectra to identify the \(m/z\) of the peptide core was problematic even when the quality of the spectra was good. Furthermore it was found that the analysis of the unfraccionated ULF samples was invariably dominated by MS/MS performed on the more abundant glycopeptides of lactotransferrin and other non-glycosylated peptides despite the PID mode used. To additionally identify protein carriers of multifucosylated \(N\)-glycans, the total digest was therefore fractionated on agarose-bound \(A. auran\textit{tia}\) lectin (AAL), which specifically recognizes terminal Fuc at various positions. MALDI MS profiling of the \(N\)-glycans released from the unbound and bound fractions confirmed that the latter contained only peptides glycosylated with complex type \(N\)-glycans with at least three Fuc residues in the context of Le\(^a\)/Le\(^\delta\) epitopes, whereas those with high mannose or a lower degree of fucosylation were collected in the non-binding fraction (data not shown). Subsequent LC-MS/MS analysis of the glycopeptides captured by the AAL after peptide-\(N\)-glycosidase F treatment led to identification of several other glycosylation sites corresponding to \(10\) other \(N\)-glycosylated. Incidentally manual MALDI MS/MS analysis of an originally unassigned \(m/z\) 2374 (Fig. 5B) afforded an abundant \(y_5\) ion due to facile cleavage at Asp and indicated that Asn\textsuperscript{65} is an \(N\)-glycosylation site. The 115.0-mass unit difference between \(y_\text{A}\) and \(y_\text{B}\) clearly showed that it is an Asp and not Asn residue. Other prominent \(y\) ions further confirmed the peptide sequence, and in particular, the three consecutive ions \(y_{10}, y_11,\) and \(y_{12}\) firmly established that Asn\textsuperscript{65} was not \(N\)-glycosylated. Incidentally manual MALDI MS/MS analysis of an originally unassigned but most intense signal at \(m/z\) 2534 (Fig. 5B) afforded its identification as the \(N\)-terminal peptide of 24p3 with a blocked \(N\)-terminal pyroglutamate residue, thus confirming a previous report of this modification (31).

**Terminal Fucosylated Epitopes Contribute Specifically to Sperm Motility-enhancing Activity of 24p3**—The emerging picture from concerted glycomics and proteomics analyses described above indicated that a variety of ULF glycoproteins including the two major components, lactotransferrin and 24p3, are carriers of Le\(^a\)/Le\(^\delta\). Because it has been shown that 24p3 can bind to sperm and enhance its motility (32), it would be of interest to see whether this binding is enabled or mediated by its unusually rich Fuc content and, if so, may be extended to other ULF glycoproteins similarly carrying the fucosylated epitope such as lactotransferrin. We first showed that both FITC-conjugated 24p3 and lactotransferrin could indeed bind to sperm by flow cytometry analysis. However, it is clear that 24p3 exhibited a better binding and was less affected by competing lactotransferrin than the other way around (Fig. 6A). Using the computer-assisted sperm assay, it was found that addition of 24p3 indeed enhanced the sperm motility, whereas lactotransferrin did not exhibit such an effect relative to control (Fig. 6B). Interestingly the motility-enhancing activity of 24p3 could be abolished by co-incubation with oligosaccharides presenting the Le\(^a\), Le\(^\delta\), and type II \(H\) (Fuc\textsubscript{α1–2}Gal\textsubscript{β1–4}GlcNAc\textsubscript{β1–}) epitopes but not by those presenting terminal LacNAc or 6\(^s\)-sialyl LacNAc (Neu5Ac\textsubscript{α2–6}Gal\textsubscript{β1–4}GlcNAc\textsubscript{β1–}). The results therefore suggest that effective binding to sperm and priming of its motility may be restricted to 24p3 and are partly mediated by its multiply fucosylated \(N\)-glycans. It is glycan-specific because its activity could only be inhibited by fucosylated terminal epitopes, whereas non-fucosylated and/or sialylated counterparts exerted no appreciable inhibitory effects under the same exper-
Glycoproteomics of Mouse Uterine Luminal Fluid
Glycoproteomics of Mouse Uterine Luminal Fluid

Identification of de-N-glycosylated A. aurantia lectin-enriched glycopeptides from ULF by nano-LC-MS/MS analysis

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Accession number</th>
<th>De-N-glycosylated peptide sequence&lt;sup&gt;a&lt;/sup&gt;</th>
<th>m/z (observed)</th>
<th>z</th>
<th>Ion score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lactotransferrin precursor</td>
<td>gi</td>
<td>21431833</td>
<td>DSSNFHLNQLQGLR</td>
<td>815.4088</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TAGWNIPMGLADOTR</td>
<td>872.4540</td>
<td>2</td>
</tr>
<tr>
<td>Mucin 4</td>
<td>gi</td>
<td>124378015</td>
<td>TDLSWLVQPK</td>
<td>593.8275</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>VVVYQDTAVAFNTR</td>
<td>763.8580</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TIGGSTSINFEQHMNDTNQYPPSIQCSSK+CAM (C)</td>
<td>1214.2365</td>
<td>3</td>
</tr>
<tr>
<td>Alkaline phosphatase</td>
<td>gi</td>
<td>22138740</td>
<td>DRTDVEYLEDEK</td>
<td>504.5649</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TRCQTTGQNEVTSILR+CAM (C)</td>
<td>617.6408</td>
<td>3</td>
</tr>
<tr>
<td>Leucine-rich α2-glycoprotein</td>
<td>gi</td>
<td>16418335</td>
<td>MFSQDTR</td>
<td>500.2060</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SLPPGFLSTSAADLSTVLTR</td>
<td>658.7048</td>
<td>3</td>
</tr>
<tr>
<td>Aminopeptidase N</td>
<td>gi</td>
<td>31077182</td>
<td>NPNLNTIHBNLR</td>
<td>468.9086</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>QVTPLFFYFQDR</td>
<td>780.9080</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KGVTLFFYFQDR</td>
<td>563.6398</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SDVTRPSTFWIYWAPFQLK</td>
<td>831.7794</td>
<td>3</td>
</tr>
<tr>
<td>CD14 antigen</td>
<td>gi</td>
<td>6753332</td>
<td>NPSDELQVGDLSLK</td>
<td>854.9357</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LDNNPSDELQVGDLSLK</td>
<td>698.3500</td>
<td>3</td>
</tr>
<tr>
<td>Complement factor B</td>
<td>gi</td>
<td>6996919</td>
<td>SPFYQLSDQISFQTYGYVLR+CAM (C)</td>
<td>1287.0997</td>
<td>2</td>
</tr>
<tr>
<td>Complement factor B</td>
<td>gi</td>
<td>110019</td>
<td>GSDQTYKPR</td>
<td>469.9103</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>gi</td>
<td>397800</td>
<td>GDLTEDTYKPR</td>
<td>469.9103</td>
<td>3</td>
</tr>
<tr>
<td>Chloride channel</td>
<td>gi</td>
<td>8567336</td>
<td>DHHFTLDVQDTTGLPQ</td>
<td>943.9733</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ASIQGGNLVDTWTAPGDDYDHGR</td>
<td>877.0939</td>
<td>3</td>
</tr>
<tr>
<td>Tissue factor precursor (TF) (CD142 antigen)</td>
<td>gi</td>
<td>135667</td>
<td>AFIDLWISTDFK</td>
<td>722.3613</td>
<td>2</td>
</tr>
<tr>
<td>Complement factor H precursor</td>
<td>gi</td>
<td>116132</td>
<td>LTFETHQSTMDYK</td>
<td>529.9050</td>
<td>3</td>
</tr>
</tbody>
</table>

<sup>a</sup> The MS/MS spectra for all de-N-glycosylated peptides identified with an ion score threshold of 15 were manually examined for Asn to Asp conversion (marked by D) at the consensus sequon (annotated MS/MS spectra are provided in supplemental Data 3). CAM (C), carbamidomethyl modification of Cys.

Fig. 4. Identification of two glycopeptides corresponding to the two N-glycosylation sites of lactotransferrin by LC-MS/MS analysis. MS survey scans summed over two different periods of elution times (A and B) show the full range of glycoforms corresponding respectively to the two unique tryptic glycopeptides of lactotransferrin. The triply charged molecular ion signals within each cluster are related by the residual mass difference of a Fuc (m/z = 48.67 for 3+) as indicated. However, two Fuc residues and one Neu5Ac differ only by 1 mass unit and therefore cannot be clearly delineated in the presence of natural isotopes. Manual examination of the MS/MS spectra of the precursors at m/z 1278.78 (C) and 1317.48 (D), representative of each of the two unique sites, revealed that both Neu5Ac-Gal-GlcNAc and Fuc-Gal-(Fuc)GlcNAc epitopes are present. The MS/MS spectra were manually interpreted and assigned as annotated. Symbols used are as in previous figures.
In direct comparison with the MALDI-TOF/TOF-MS/MS spectrum of the same sodiated molecular ion at m/z 2592 from ULF (Fig. 2A), the corresponding trifucosylated biantennary complex type N-glycan structure from SVF clearly afforded only fragment ions that are indicative of Lex but not Ley epitope (Fig. 8A). Thus, the non-reducing terminal ions specific to Ley, in particular the sodiated C ion at m/z 433 and the B and E ion pair at m/z 843 and 763, were absent. Likewise the D ion at m/z 1228 and 3,5A ion at m/z 1126 were not detected, indicative of the absence of a difucosylated LacNAc antenna on the 6-arm. Instead the much simpler MS/MS spectrum is fully consistent with a rather homogeneous core fucosylated biantennary structure with a single Lex on both antenna as supported by the characteristic D ion at m/z 472 (30). In addition, under low energy CID MS/MS on MALDI-Q-TOF instrument (Fig. 8B), consecutive losses of both monofucosylated antennae produced a trimannosyl core ion at m/z 1317 (and m/z 866) with two free OH groups specific to a biantennary N-glycan structure (29).

Interestingly, similar low energy CID MS/MS on the sodiated molecular ion at m/z 3215, which contained an additional Fuc1Hex1(HexNAc)1 unit, afforded the same trimannosyl core pair at m/z 1317 and 866 (Fig. 8C) and therefore likewise established it as a biantennary structure. The presence of a dimeric Lex on one antenna was also evident from the B ion at m/z 1283 coupled with a direct loss of such a difucosylated antenna from the parent to give the Y ion at m/z 1954. Neither this nor the corresponding high energy CID MS/MS (data not shown) on the tetrafucosylated biantennary structure could detect any fragment ion specific to Ley. Thus, in comparison with ULF, the glycome of SVF is clearly devoid of Ley but does share the common feature of Lex as a prominent terminal epitope along with sialylated LacNAc. In contrast, the glycome of EF does not contain any appreciable amount of fucosylation, and the complexity of N-glycan processing increases from the region proximal to testis to the distal end.

**DISCUSSION**

Despite recent technical advances in both MS instruments and front end sample preparation, glycoproteomics analysis remains daunting. In principle, all requisite information for
unambiguous identification of the protein carrier, its glycosylation site(s), and the exact glycan structures encompassing a full range of heterogeneous glycoforms can be obtained from detailed MS/MS analysis of the glycopeptides. In practice, however, this premise is limited by detecting the glycopeptides in the first place and the subsequent difficulties in obtaining a full range of sequence-informative MS/MS fragment ions pertaining to both the peptide and glycan sequence at sufficiently high sensitivity. In particular, glycan sequencing that would unambiguously determine its linkage-specific substituents and branching pattern is usually not feasible by MS/MS analysis of the total tryptic digests. Instead it is more readily accomplished by analyzing the permethylated derivatives of the released glycans either by ion trap-based MS or, as in our case, by complementary low and high energy MALDI CID MS/MS (29, 30, 34). With additional exoglycosidase digestion and GC-MS analysis, we have unambiguously defined the ULF glycome as comprising core fucosylated complex type N-glycans with multiple Le\(^\alpha\) and Le\(^\beta\) epitopes to account for its high Fuc content along with a variable amount of Neu5Aca2–6-sialylated antenna. This overall glycomic pattern is reflective of its constituent glycoproteins including the two most abundant ones, 24p3 and lactotransferrin.

Because 24p3 was purified to homogeneity and N-glycans were released thereof, attribution of the glycosylation profile was not an issue. We have additionally found that only one of two potential sites was N-glycosylated while confirming that its N terminus was indeed modified by pyroglutamation. In the case of lactotransferrin, direct LC-MS/MS analysis of the glycopeptides in conjunction with identification of the de-N-glycosylated peptides by MALDI MS/MS have collectively verified the two occupied N-glycosylation sites, both carrying the implicated N-glycan structures. However, identification of other glycopeptides was less conclusive. We have to resort instead to a commonly used approach, namely to first isolate the highly fucosylated glycopeptide pool by lectin affinity and subsequently identify the proteins based on MS/MS analysis of the de-N-glycosylated peptides with a requisite Asn to Asp conversion. Not all of the 10 ULF proteins thus identified were in the list of the identified ULF proteins by shotgun analysis of the total tryptic digests. This would be expected because the lectin affinity capture would give an enrichment effect for some of the less abundant proteins carrying highly fucosylated N-glycan structures. Glycomics analysis of this lectin-captured pool of glycopeptides indicated that all N-glycan structures released thereof contained at least three Fuc residues, thus containing a minimum of one terminal Lewis epitope in addition to core fucosylation. This largely rids the identified de-N-glycosylated peptides of false positives due to nonspecific bindings commonly associated with affinity capture.

Conversely not all abundant ULF proteins identified by shotgun proteomics analysis appeared to carry the multifucosylated N-glycan structures, although false negatives could be contributed by many factors including random misses. By high stringency statistical filtering, we reported in this work 44 ULF proteins deemed identified with high confidence. A problem with analysis of the ULF proteome, as in the case of plasma, was the predominant abundance of a few major components such as complement C3 and lactotransferrin, which precluded more proteins to be identified by a single dimensional shotgun analysis of non-fractionated samples despite the use of a high performance instrument. An initial attempt at two-dimensional gel separation followed by spot-wise identification also did not result in a more comprehensive coverage of the ULF proteome. Likewise the analysis of the
glycopeptides was dominated by repeated MS/MS on the glycoforms of lactotransferrin. Because the main focus of this work was centered on detailed glycomics analysis and to obtain a simplistic overview of the prevalent characteristics of the ULF glycoproteome, we have not delved further into the depth of proteomic coverage.

For the more abundant proteins, the three major groupings according to the apparent biological processes in which they

Fig. 7. Comparative glycomics mapping of mouse EF from the caput (A), corpus (B), and cauda (C) regions and SVF (D). MALDI MS analyses were performed on the permethylated N-glycans derived from tryptic digested total EF and SVF proteins similar to the analyses of ULF glycome (Fig. 1A) for direct comparison. Spectra are shown in two overlapping segments normalized to the intensities of respective base peaks. The approximate -fold of magnification from the first to second segments is indicated. In C, sialylation on the hybrid structure appears to be restricted to Neu5Ac (○), whereas the mass increments are consistent with Neu5Gc (△) sialylation on the complex type structures. In D, signals corresponding to the major tri- and tetra-fucosylated structures were further subjected to MALDI MS/MS analyses (Fig. 8) to confirm structural assignment. The multifucosylated complex type structures are accompanied by satellite peaks at multiples of 30 mass units higher, which may be contributed by terminal α3-Gal instead of fucosylation but were not further confirmed by MS/MS analyses. Symbols used are as defined in Fig. 1. Comparative SDS-PAGE profiles of the ULF, EF, and SVF, visualized either by SYPRO Ruby stain (left panel) or AAL and L. tetragonolobus lectin (LTL) blots (right panel) are also shown as an inset.
may be involved are: 1) immune response, including proteins involved in complement pathways such as complement C3, complement factor H, complement factor B, and modulators involved in the innate immune response such as CD14 anti-

Fig. 8. MALDI MS/MS sequencing of tri- and tetrafucosylated complex type N-glycans from SVF. MS/MS analyses were performed on both MALDI-TOF/TOF and MALDI-Q-TOF instruments for the major multifucosylated peaks detected by MALDI MS profiling of the permethyl derivatives (Fig. 7D). Shown representatively here are the respective high (A) and low energy (B) CID MS/MS spectra for the trifucosylated precursor at m/z 2593 and the corresponding low energy CID MS/MS spectra (C) for the tetrafucosylated precursor at m/z 3217. Annotation symbols and ion nomenclature are as defined in Fig. 2. Fragment ion peaks not annotated for ion type in B and C correspond to multiple cleavages. Ions annotated as further loss of the reducing terminal Fuc-HexNAc may be viewed alternatively as arising from further loss of non-reducing terminal Le" units from the B ion formed at the chitobiose core (m/z 2141 and 2765, respectively). Elimination of a Fuc from the parent ion as annotated is obvious and indicative of a 3-linked fucosylation on HexNAc as in Le".

gen, lipopolysaccharide (LPS)-binding protein, and peptidoglycan recognition protein; 2) metabolism, including several enzymes such as aminopeptidase N, α 3A chain of laminin-5, alkaline phosphatase, chitinase 3-like protein 1, and cathep-
An abundance of immune response-related proteins detected in ULF is not surprising because a reported function of the female reproductive tract fluid is to protect against infection via the innate and adaptive immune systems (35). Except for proteins directly involved in the complement pathway, peptidoglycan recognition protein and LPS-binding protein are known as pathogen-associated molecular patterns receptors, and CD14 mediates the response to LPS. Even lactotransferrin has been accorded an important role in preventing infection from pathogens in human cervical-vaginal fluid (36). Lactotransferrin is also known to interact with accessory molecules involved in the TLR4 pathway, including CD14 and LPS-binding protein, suggesting that it may activate components of the TLR4 pathway (37, 38). Likewise several of the protease or antiprotease proteins identified have been detected in human uterine and cervical-vaginal fluid (36, 39). Among these proteases, aminopeptidase N (CD13), which is a type II metalloprotease, is involved in the trimming of antigen and the process of antigen presentation (40). Cathepsin B is a cysteine protease that functions in protein catabolism and may also be involved in processing of antigens in the immune response as well as hormone activation (41). Other proteins such as laminin play a central role in the formation and stability of basement membrane, and chitinase 3-like protein is associated with down-regulating the catabolic or degradative aspects of the inflammatory response (42).

More intriguingly is the elusive function of 24p3, a major component of ULF. The 24p3 protein has been implicated in diverse physiological processes, including apoptosis, iron transport, proinflammatory processes, and cell survival (43-46). Notably in the context of the reproductive system, functional interactions with spermatozoa have been indirectly demonstrated through its sperm motility-enhancing and acrosome reaction-suppressing activities (32, 47) as well as its internalization for ferric ion delivery (48). Here we provided evidence that 24p3 could directly bind to spermatozoa and showed by sperm motility assay that co-incubation of 24p3 with oligosaccharides containing specific fucosylated terminal epitopes in the forms of Lea, Leb, and type II H, but not terminal LacNAC or 6'-sialyl LacNAC, would attenuate its enhancer activity. However, it seems that not all ULF proteins carrying the same fucosylated epitope would bind and interact similarly. In particular, only 24p3, but not lactotransferrin, could visibly enhance the sperm motility. One hypothesis is that the relative density and multivalent presentation of the implicated glycoepitopes may be the critical determining factors. Structural modeling suggested that the N-glycosylation sites of these two proteins were similarly exposed on the surface (data not shown). However, the relatively smaller size of 24p3 may contribute to a better contact with the spermatozoa by exhibiting a higher glycan density relative to the larger lactotransferrin where the two occupied N-glycosylation sites are individually located on each of the two lobes. Another hypothesis is that not only the glycan moiety but the protein structure and conformation itself are also essential for productive interaction. To distinguish between these possibilities, further assays using recombinant 24p3 expressing no or other types of glycan structures will be helpful.

Irrespective of the putative roles of protein carriers, the biological functions of Lea per se in mouse reproductive system remain unclear. Interactions with spermatozoa would implicate the presence of a specific lectin on the sperm surface that has yet to be identified. Taking a different approach, our comparative glycomics analyses of mouse reproductive luminal fluids from male to female, including EF, SVF, and ULF, clearly indicated that the expression of Lea, which requires the activity of α2-fucosyltransferase, was restricted to female ULF, whereas Leb was not found in the epididymal tract but was present in SVF and ULF. Previous study has further demonstrated that the expression of the α2-fucosyltransferase mRNA transcript in the uterine epithelium cells is dynamically regulated in concert with the estrous cycle (49). In this work, the ULF sample was collected from 3-week-old female mice pretreated with DES, which is a synthetic estrogen that acts via the same mechanism as estrogen in stimulating the mouse uterine epithelial secretory response. DES binds to intracellular estrogen receptors with an affinity ~100 times higher than the endogenous estrogen (50). Treatment of mice with DES would markedly increase the production of ULF, thus providing a more readily available sample source for our detailed analyses. To evaluate whether this treatment truly reflected the physiological situation, we additionally collected and analyzed the ULF samples from female mice at the estrous phases of the cycle. 1D SDS-PAGE analysis showed a protein expression pattern similar to that from DES-treated sample except for less extensive complement C3 degradation, whereas Western blot analysis against Leab likewise indicated no obvious difference in its expression pattern (data not shown). More importantly, when comparative MS glycomics mapping was performed for the N-glycans derived from different stages of the estrous cycle, a high expression level of Leab epitopes was only detected during the estrous phase with a much a lower level during proestrous phase and a level that was hardly detectable during metestrous and diestrous phases (data not shown).

Thus, both Leab and 24p3 expression levels in ULF are elevated during the estradiol surge in proestrous and estrous phases and suppressed when progesterone rises to a high level at metestrous and diestrous phases. Its sperm-binding activity and motility-enhancing properties therefore correlate with the fertile window of mouse estrous cycle. This is in
contrast to human glycolelin, which carries Le" specifically but in the glycoform found in the male seminal plasma but not that from female uterine origin (11). The expression of the latter glycoform was found to be at a minimum during the periovulatory midcycle but increased significantly from the 4th postovulatory day and contributed to the contraceptive activity in the latter half of the secretory phase of the menstrual cycle. Thus, hormone-regulated, gender-specific glycosylation is a common feature for human and mouse reproductive tracts. Both systems involve Le" specifically but in a different context. A unifying model for functional roles at the molecular level is still lacking at present but would surely benefit from current glycobiology and glycoproteomics studies.

Acknowledgments—LTQ-Orbitrap data were acquired at the Academia Sinica common mass spectrometry facilities located at the Institute of Biological Chemistry. We are grateful to He-Hsuan Hsiao for technical assistance in the early stage of this work and Dr. Han-Jia Lin for providing the SVF samples for analysis.

* This work was supported by Academia Sinica and Taiwan National Science Council Grants 92-2311-B-001-091 (to K. H. K.) and 95-3112-B-001-014 and 96-3112-B-001-018 (to the National Research Program for Genomic Medicine Core Facilities). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** To whom correspondence may be addressed: Inst. of Biological Chemistry, Academia Sinica, 128 Academia Rd. Sec 2, Nan-kang, Taipei 11529, Taiwan. Fax: 886-2-27889759; E-mail: stc316@gate.sinica.edu.tw.

† To whom correspondence may be addressed: Inst. of Biological Chemistry, Academia Sinica, 128 Academia Rd. Sec 2, Nan-kang, Taipei 11529, Taiwan. Fax: 886-2-27889759; E-mail: khkoo@gate.sinica.edu.tw.

REFERENCES

adenocarcinoma by tandem mass spectrometry sequencing analyses. J. Biol. Chem. 283, 16455–16468