Regulation of Glycan Structures in Animal Tissues

TRANSCRIPT PROFILING OF GLYCAN-RELATED GENES

Alison V. Nairn, William S. York, Kyle Harris, Erica M. Hall, J. Michael Pierce, and Kelley W. Moremen

From the Complex Carbohydrate Research Center and the Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602

Glycan structures covalently attached to proteins and lipids play numerous roles in mammalian cells, including protein folding, targeting, recognition, and adhesion at the molecular or cellular level. Regulating the abundance of glycan structures on cellular glycoproteins and glycolipids is a complex process that depends on numerous factors. Most models for glycan regulation hypothesize that transcriptional control of the enzymes involved in glycan synthesis, modification, and catabolism determines glycan abundance and diversity. However, few broad-based studies have examined correlations between glycan structures and transcripts encoding the relevant biosynthetic and catabolic enzymes. Low transcript abundance for many glycan-related genes has hampered broad-based transcript profiling for comparison with glycan structural data. In an effort to facilitate comparison with glycan structural data and to identify the molecular basis of alterations in glycan structures, we have developed a medium-throughput quantitative real time reverse transcriptase-PCR platform for the analysis of transcripts encoding glycan-related enzymes and proteins in mouse tissues and cells. The method employs a comprehensive list of >700 genes, including enzymes involved in sugar-nucleotide biosynthesis, transporters, glycan extension, modification, recognition, catabolism, and numerous glycosylated core proteins. Comparison with parallel microarray analyses indicates a significantly greater sensitivity and dynamic range for our quantitative real time reverse transcriptase-PCR approach, particularly for the numerous low abundance glycan-related enzymes. Mapping of the genes and transcript levels to their respective biosynthetic pathway steps allowed a comparison with glycan structural data and provides support for a model where many, but not all, changes in glycan abundance result from alterations in transcript expression of corresponding biosynthetic enzymes.

Carbohydrate structures attached to glycoproteins, glycolipids, and proteoglycans have been shown to play key roles in a variety of biological recognition events (1). Although there are many examples of the contribution of N-glycans to the bioactivity, folding, localization, and immunogenicity of the attached polypeptide, the functional roles of individual oligosaccharide structures on a given glycoprotein are difficult to predict (1–5). At the cellular level, N-linked, O-linked, and glycolipid glycan structures have been shown to contribute to several essential aspects of biological recognition, including cell adhesion during development, immune surveillance, inflammatory reactions, hormone action, viral infection, arthritis, and metastasis of oncogenically transformed cells (6–9). Most of our understanding of the roles of cellular glycosylation in physiology and pathology comes from a combination of glycan structural analysis on specific glycoproteins, cell surfaces, or total tissue extracts in combination with years of study on the biochemistry and enzymology of glycan biosynthetic and degradative enzymes (10–13). Despite this array of biochemical and genetic information, very little is known about the global regulation of glycan synthesis and degradation.

A major goal in the field of “glycobiology” is an understanding of how glycan structures are regulated in abundance and the impact that these changes have on the physiology and pathology of an organism. Several difficulties arise when attempting to examine the regulation of glycan structures in complex biological systems. Because glycan biosynthesis is a post-translational modification, it is not directly template-driven like the synthesis of polypeptide structures from genome-derived transcripts. Thus, numerous factors can impact the efficiency and penetration of individual glycosylation steps on protein and lipid acceptors, including enzyme accessibility to glycan modification sites, the abundance of the respective protein or lipid acceptors, availability of sugar-nucleotide precursors, and relative enzyme levels or relative localization of biosynthetic enzymes that can compete for the same glycan substrates. Despite these complexities in glycan biosynthesis, several lines of evidence indicate that one of the major modes of regulating cellular glycosylation is transcriptional regulation of the enzymes involved in glycan synthesis and catabolism (14). One method for testing whether the elaboration of glycan structures is controlled at the transcriptional level is by the comparison of glycan structural data with transcript abundance measurements in multiple biological samples, where differences in glycan structures are known to occur. The last decade has seen significant advancements in methods for glycan structural analysis providing increased breadth, depth, and sensitivity to the glycan structures detected and quantitated within a single experiment (15–17). Although these analyses have revealed critical changes in glycan structures during development or between biological samples, rarely have they been paired with broad-based transcript analysis to determine whether tran-
scriptional regulation is the major mechanism driving the structural alterations (14, 18–24).

Transcript profiling of glycan-related genes has its own set of complexities. The enzymes involved in glycan synthesis and modification have been collated into multifamily genes based on sequence and structural similarities. In mammalian cells, glycosyltransferases number ~200 members and are subdivided into 40 families (CAZy database (25, 26)), but in many cases the acceptor specificity of individual family members is not known or potential enzymatic redundancy may exist between multiple members of the same enzyme family. Thus, one-to-one mapping of individual gene products to steps in glycan biosynthetic pathways is difficult to achieve or may have ambiguity among multiple family members. Existing web-based resources (CAZy (25, 26), KEGG (27–29), Consortium for Functional Glycomics (14, 18, 21), and SOURCE (30)) have collated and annotated many of the genes related to glycan biosynthesis, but comprehensive resources for mapping enzymes to complex glycan biosynthetic pathways for glycoprotein, glycolipid, and proteoglycan biosynthesis and catabolism are still in their early stages.

An additional complexity for the study of glycan-related gene expression is the relatively low abundance of transcripts encoding many of the critical enzymes involved in glycan modifications. These low transcript levels make it difficult to employ broad-based survey methods, such as microarray approaches (14, 18–22, 24), for global transcriptome analyses. More focused approaches employing quantitative real time PCR (qRT-PCR)2 have been employed effectively (23, 31, 32), but this strategy has generally been restricted to the analysis of a relatively small number of target genes.

We have chosen to develop a broad-based analytical platform for transcript analysis of glycan-related genes that has three key components. First, we have drawn on numerous publicly available resources and the primary literature to generate a comprehensive gene list encoding enzymes and proteins involved in glycobiochemistry, including sugar-nucleotide biosynthesis, transporters, glycan extension, modification, recognition, catabolism, and numerous glycosylated core proteins (>700 genes in the mouse). Second, we have developed a robust, sensitive, and flexible qRT-PCR platform for transcript analysis using experimentally validated primer sets for all of the members of the mouse gene list. This strategy has allowed us to examine global changes in glycan-related transcripts for both highly expressed genes as well as low abundance transcripts that may play key roles in generating important glycan epitopes. Third, we have developed a set of detailed pathway diagrams for glycan biosynthesis and modification and initiated the mapping of all members of the gene list to their respective biochemical pathway steps. Initial use of these pathway diagrams has allowed the visual depiction of transcript abundance within a framework of glycan biosynthetic pathways as a means of correlating glycan structural data with transcript abundance.

2 The abbreviations used are: qRT-PCR, quantitative real time polymerase chain reaction; MALDI-MS, matrix-assisted laser desorption ionization mass spectrometry; RMA, robust multichip average; qDNA, genomic DNA; Rpl4, ribosomal protein L4; EST, expressed sequence tag; IGnt, I-GlcNAc transferase; GPI, glycosylphosphatidylinositol; NCAM, neural cell adhesion molecule.

Transcript Profiling of Mouse Glycan-related Genes

As an experimental framework for examining the regulation of glycan structures in mammalian systems, we have analyzed RNA samples derived from several adult mouse tissues and compared them with microarray data from parallel tissue samples and glycan structural data previously obtained by MALDI-MS approaches (14, 16, 17). Greater sensitivity and dynamic range was found for our qRT-PCR approach compared with focused microarrays, particularly for the numerous low abundance glycan-related enzymes. Comparison with glycan structural data demonstrated numerous correlations between glycan structures and transcript abundance for their respective biosynthetic enzymes. Several cases were also noted where differences in glycan structures did not correlate with transcript abundance suggesting that regulation may occur at a post-transcriptional level. The analysis of glycan-related transcripts within the context of biosynthetic pathways also predicted differences in low abundance glycans consistent with previous observations of these structures in the literature.

EXPERIMENTAL PROCEDURES

Compilation of the Glycan-related Gene List—Our murine glycan-related gene list was compiled from several sources, including the following: the database of Carbohydrate Active Enzymes (26), a web-based genomic resource for animal lectins (33) organized by Dr. Kurt Drikamer, the Kyoto Encyclopedia of Genes and Genomes (27–29); the gene list for the GLYCOv2 Gene Chip from the Consortium for Functional Genomics (14, 18, 21); the microarray gene list from the Glyco-Chain Expression Laboratory (20, 22); the Transport Classification database (34); NCBI (www.ncbi.nlm.nih.gov (35)); SOURCE (30); contributions from collaborating investigators, and extensive searches of the primary literature (see Table 1 for gene list categories, member totals, and sources as well as supplemental Table 1 for the detailed gene list, including additional gene annotation information). Unique NCBI gene identifiers (GeneIDs) (36) for each member of the gene list were used to check for isoforms of a single gene to prevent duplications of gene entries in the list. Members of gene families that were >95% identical at the DNA sequence level were treated as the same gene, and one primer pair was designed to amplify a region that was 100% identical in sequence. Genes that encode proteins with multiple functions (i.e. glycosyltransferase activity and carbohydrate binding domains) were grouped by their catalytic activity to prevent redundant entries for primer design. Genes that encode proteins with two separate catalytic activities were listed under both categories in the gene list, but only one primer pair was designed for the transcript.

Primer Design—Primer pairs were designed for a number of housekeeping genes as well as the glycan-related genes using a restricted set of conditions. Coding region sequences for a specific gene were compared with the corresponding genomic sequence in the NCBI database via the BLAST search algorithm (37) to determine intron/exon boundaries. A single coding exon was submitted to the Primer3 web-based primer design program (38) with the following parameters: product size ranges, 65–75 bp; primer size, 19–21 bp; primer Tm, 59–61 °C with a maximum Tm difference between primers for a given gene of 1 °C, maximum self-complementarity of 6 bases, maxi-
Transcript Profiling of Mouse Glycan-related Genes

uum 3′ self-complementarity of 5 bases, and maximum repeat of a single base (poly(X)) of 5 bases. All other settings were the default values. Primers were synthesized by MWG Biotec (High Point, NC).

Primer Validation—Amplification reactions consisted of 5 μl of diluted mouse genomic DNA (a kind gift from Dr. Nancy Manley, Department of Genetics, University of Georgia) as template, 5 μl of primer pair mix (500 nm each primer, 125 nm final concentration) (MWG Biotec), and 10 μl of iQ™SYBR® Green Supermix (Bio-Rad). Amplifications were performed in a 96-well iCycler or myIQ real time detection system (Bio-Rad) with the following cycling conditions: 95 °C for 3 min, followed by 40 cycles of: 95 °C for 10 s (denaturing), 65 °C for 45 s (annealing), 78 °C for 20 s (data collection), followed by a melt curve program (95 °C for 1 min and 55 °C for 1 min and then increasing temperature of 0.5 °C per cycle for 80 cycles of 10 s each). We found that collecting fluorescence data at a temperature above the annealing temperature yielded cleaner amplification profiles, because any primer dimers that formed would be dissociated at this temperature as has been described previously (39). Primer pairs were tested at a single DNA concentration in triplicate, and the average of the cycle threshold (Ct) values was compared with that of a housekeeping gene (Rpl4). Primer pairs that yielded an average Ct within 2 units of the average Ct for the control gene were tested for efficiency, and those outside the 2 Ct window were re-designed (supplemental Fig. 1A). A typical amplification curve from a genomic DNA dilution series is shown in supplemental Fig. 1B. The efficiency of amplification for each primer pair was determined in duplicate using serial dilutions of mouse genomic DNA as the template by the method of Liu and Saint (40). The Standard Curve Method (41) was applied to the analysis of data from each primer set to generate plots of Ct versus log concentration of template and the slope was used to determine amplification efficiency, where efficiency (E) = 10−1/slope – 1 (supplemental Fig. 1C). For validation purposes, we selected an acceptable range of 100 ± 10% efficiency with genomic DNA as template (shown as dotted lines in supplemental Fig. 1C). Following the amplification and melt curve analysis, data were set to a common threshold, and the efficiency of the primer pair was determined from the slope of the standard curve using software supplied with the qRT-PCR instrumentation (Bio-Rad). Melt curves were analyzed for the presence of a single peak of −d(RFU)/dT at 80–86 °C (where RFU is relative fluorescence units) indicating a single amplification product (supplemental Fig. 1D). An example of a melt curve analysis where a primer pair amplified more than one product is shown in supplemental Fig. 1E. Primers that failed any of the validation steps were redesigned and reanalyzed until a suitable primer pair was obtained. Our success rate for primer design in our first attempt was ~90%.

RNA and mRNA Isolation—Kidney, liver, testis, and brain tissue were isolated from C57BL6 mice (a kind gift from Dr. Mary Bedell, University of Georgia), flash-frozen in liquid N2, and stored at −80 °C. Total RNA was isolated using TRIzol reagent (Invitrogen) according to the manufacturer’s instructions. Samples were digested with RNase-free DNase I (Ambion) to remove genomic DNA contamination and then re-extracted with TRIzol, precipitated with isopropyl alcohol, resuspended in diethyl pyrocarbonate-treated water, quantitated, and stored at −80 °C. Poly(A+) mRNA was isolated from total RNA using Dynabeads mRNA direct kit (Dynal, Invitrogen), quantitated, and stored at −80 °C.

cDNA Synthesis—cDNA was synthesized from 500 ng of poly(A+) mRNA using the SuperScript III First Strand synthesis kit (Invitrogen) according to the manufacturer’s instructions except that both oligo(dT) and random primers (1:1) were included in the cDNA synthesis reactions. A control reaction lacking reverse transcriptase (“No-RT”) was prepared and analyzed to detect the presence of contaminating genomic DNA. For qRT-PCR reactions, cDNA reaction products (20 μl) were diluted 1:20 in water and used as template in triplicate reactions for each primer pair assayed.

Normalization Gene Selection—qRT-PCR reactions with cDNA templates from mouse tissues were assayed using several housekeeping genes to determine the variability of expression across all tissues. The gene with the lowest variation across all tissues was selected as the normalization gene for all samples.

RT-PCR Data Analysis—Reactions for qRT-PCR were set up in a 96-well plate format using the same amplification conditions described above for primer validation using genomic DNA as template. The No-RT control cDNA template was tested with several primer pairs to confirm that the sample was free of contaminating genomic DNA prior to analysis of the reverse-transcribed template. Each primer pair was analyzed in triplicate for each cDNA sample. Following each run, the threshold was set to a common value to maintain consistency between runs, and data for each primer pair was averaged and the standard deviation determined. We chose an arbitrary cutoff of 0.5 Ct for the standard deviation (42). Triplicate values with a standard deviation >0.5 Ct were re-assayed. The raw fluorescence data from the PCR machines were also analyzed using LinRegPCR (43) to determine the amplification efficiency of the individual reactions, and a cutoff of <5% was set as acceptable variability. Averaged Ct data were transformed to linear amplimer abundance values (2−Ct) and normalized to the housekeeping gene (Rpl4).

Data Analysis Method—We utilized the ΔΔCt method (41) to determine the relative transcript levels for the glycan-related genes in a given cDNA sample. This analysis method requires the assumption that the amplification efficiencies of all reactions are approximately equal. A test for equal efficiencies is to plot ΔCt (Ctgene − Ctcontrol) versus template concentration for a dilutions series and ensure that the slope of the generated line is <0.1. Modeling conditions within these restrictions translate into an acceptable difference in amplification efficiency of <5%. Primer efficiency values from all samples tested were below the 5% cutoff (generally <3%). The normalization gene, Rpl4, was included on all 96-well plates to control for interplate and machine variations.

Microarray Analysis—Matched samples for mouse kidney and liver were analyzed on the GLYCOv2 custom microarray chip and by qRT-PCR analysis. The GLYCOv2 gene chip was produced by Affymetrix (Affymetrix, Santa Clara, CA) for the Consortium for Functional Glycomics. Samples were labeled and hybridized to the GLYCOv2 chips in triplicate as described...
Transcript Profiling of Mouse Glycan-related Genes

Organization of the mouse glycan-related gene list

A list of all glycan-related genes was compiled from several sources, as indicated in the Source column, and collated into groups and families based on enzyme or protein function and sequence. The full gene list, organized by the gene list prefix and containing additional annotation information, can be found in supplemental Table 1. Totals for each category of glycan-related genes are indicated by No. of members.

### TABLE 1

| Group | Family or subgroup | Gene list prefix | No. of members | Source |
|-------|-------------------|------------------|----------------|--------|
| Glycosyltransferases | CAZy families: GT 1–4, 6–8, 10–14, 17, 18, 21–25, 27, 29, 31–33, 35, 39, 41, 43, 47, 49–51, 54, 57–59, 61, 64–66, 68, 76 | GT | 190 | A, H |
| Glycosylhydrolases | CAZy families: GH 1, 2, 13, 18, 20, 22, 23, 27, 29–31, 33, 35, 37–39, 47, 56, 59, 63, 65, 79, 84, 85, 89, 99 | GH | 84 | A, H |
| Carbohydrate-binding modules | CAZy families: CBM 13, 20, 21 | CMB | 10 | A, H |
| Carbohydrate esterases | CAZy families: CE 1, 9, 10, 13 | CE | 21 | A, H |
| Lectins | C-type LCa | LCa | 92 | C, H |
| | I-type Lix | Lix | 16 | C, H |
| | Calnexin-like LCNX/LCRT | 2 | C, H |
| | S-type LSG | 11 | C, H |
| | P-type LPR | 2 | C, H |
| | L-type LL | 4 | C, H |
| | M-type GH47w | 7 | C, H |
| | F-box LFBX | 1 | C, H |
| Miscellaneous glycans | GH18, LAx, LMHC, LGB | 21 | D |
| Carbohydrate synthesis and metabolism | Nucleotide synthesis | NSN | 42 | E, H |
| | Transporters | TR | 28 | F, H |
| Sulfate-related | Metabolic enzymes | CM | 24 | E, H |
| | Arylsulfatases | SAS | 4 | B, E, H |
| | Sulfatases | ST | 4 | B, E, H |
| | Sulfohydrolases | SH | 1 | B, E, H |
| | Sulfotransferases | STR | 36 | B, E, H |
| Glycolipid-related | Core proteins | GL | 34 | G, H |
| Glycosaminoglycan-related | Fibroblast growth factors | FGF | 20 | B, H |
| | Fibroblast growth factor receptors | FGFR | 4 | B, H |
| Lipid-linked oligosaccharide pathway-related | | | |
| | GPI anchor biosynthesis-related | GPI | 19 | B, H |
| | Miscellaneous | MISC | 8 | B, H |
| | Housekeeping genes | | 23 | B, H |
| | | Total | 745 | |

* Sources are identified by letters A–H. A was compiled from the database of Carbohydrate Active Enzymes (25, 26). B was compiled from gene list of the GLYCO2.0 gene chip from the Consortium for Functional Glycomics (14, 18, 21) and the microarray gene list from the Glyco-Chain Expression Laboratory (20, 22). C was compiled from the web-based genomic resource for animal lectins (33). D was from K. Drickamer, personal communication. E was compiled from the Kyoto Encyclopedia of Genes and Genomes (27–29). F was compiled from the Transport Classification Database (34). G was from gene lists provided from collaborative investigators. H was compiled and edited based on database information at NCBI (www.ncbi.nlm.nih.gov (35)) SOURCE (30), and the primary literature.

* CAZy entries >90% identical at the nucleotide level were combined into a single entry in the gene list.

* Carbohydrate-binding module genes with transferase or hydrolase activities are listed as GTs or GHs to avoid redundancy.

* Within this family numerous subgroups are distinguished within the full gene list and distinctive prefixes have been appended for each subgroup as indicated by the “x.” See supplemental Table 1 for details.

* Data were also classified as GH family members, not included in total gene number.

Previously (14). The GeneChip® operating software (GCOS, Affymetrix, Santa Clara, CA) was used to determine whether genes were called “present” or “absent.” Genes were classified as present if at least two of the three calls were “present.” Robust multichip average (RMA) data, a normalized intensity value based on the amount of probe that hybridizes to the array, was used to produce signals for each gene (44), and these values were averaged and used for comparison with data generated by qRT-PCR.

**Comparison of Microarray and qRT-PCR Data**—Genes from the microarray analysis were grouped into present and absent categories. The average RMA values for each gene were plotted against the normalized linear amplimer abundance value for each gene from the qRT-PCR analysis with both data sets plotted on a log10 scale. The correlation coefficient was determined for genes that were detected as present on both the microarray and qRT-PCR analysis. The lower limit of detection for qRT-PCR was set at Ct = 35.

**Processing of Mass Spectrometry Data to Calculate Semi-quantitative Relative Glycan Abundance**—MALDI-MS data for N-glycan structures identified in samples from four adult mouse tissues were obtained from the Consortium for Functional Glycomics web site based on previously published methods (14) employing release by peptide-N-glycanase F digestion followed by permethylation. Raw spectral data posted for each mouse tissue were initially analyzed using Voyager software (Applied Biosystems, Foster City, CA) to generate lists of mean-centered ion clusters for each glycan mass followed by integration of the corresponding ion currents. In-house software was used to identify glycosyl compositions corresponding to each mean-centered ion cluster in the MS data, and the most likely glycan structure corresponding to each composition was assigned manually. Spectral features having an ion current integral of less than 1.5% of the total integrated ion current for all assigned glycan structures in the given tissue data set were not considered further. Ion current integrals for a given glycan structure were normalized and expressed as a percentage of the total ion current for the given data set.

**RESULTS**

Assembly of the Glycan-related Gene List—In an effort to examine the global regulation of glycan abundance in animal...
Transcript Profiling of Mouse Glycan-related Genes

systems, we needed to develop a medium-throughput platform for the analysis of a comprehensive “glycobiology-related” gene list that comprises all known proteins involved in the synthesis, modification, recognition, and degradation of glycoconjugates in mammalian systems. This gene list was further extended by the inclusion of numerous glycosylated core proteins, sugar-nucleotide biosynthetic enzymes, and sugar transporters. For assembly of the gene list, we drew on numerous web-based resources, contributions from collaborating investigators, and extensive searches of the primary literature (see “Experimental Procedures” and Table 1 for sources, gene list categories, and member totals). Although the initial list focused on murine genes, we recently extended the full list to human glycan-related genes (not shown), and additional genes were identified by cross-searching members from one species gene list against the corresponding other species list. The complete list of genes, including our index number, NCBI accession numbers, gene identification numbers, annotations, and validated primer sequences for each gene, are contained in supplemental Table 1.

Development of a qRT-PCR Platform for Glycobiology-related Transcript Analysis—Although numerous strategies are available for transcript quantitation in biological systems (i.e. SAGE (45–47), microarray techniques (48–50), or variations of these approaches (51, 52)), we chose to employ a qRT-PCR-based platform because of the extreme sensitivity and wide dynamic range of this latter approach (42) and its common use in validating transcript changes initially identified by microarray and SAGE techniques (53). We chose the SYBR green intercalating dye methodology to take advantage of reduced cost and ease of use in high throughput qRT-PCR compared with other approaches that employ additional primers and/or probes (54). Several considerations have been taken into account during the development of the methodology to ensure uniformity of analysis across the hundreds of target genes on the list, including standardized primer design, amplification, and validation protocols. The details of the analytical platform are described under “Experimental Procedures,” but the rationale and criteria for protocols employed are discussed below.

Validation of Amplification Efficiency and Specificity—A key component in the development of our qRT-PCR platform was the generation of primer sets that were each experimentally validated to provide uniform amplification efficiency across the entire gene list. Because cloned versions of each member of the gene list are not available for validating the amplification efficiency for a given primer set, we used genomic DNA (gDNA) as an amplification template during the extensive primer validation process described under “Experimental Procedures.” We confirmed linearity of the qRT-PCR responses across the template dilution series, calculated amplification efficiencies relative to a standard housekeeping gene (Rpl4), checked for the production of single amplimer products by melt curve analysis, and the lack of amplimer production in no template control reactions (supplemental Fig. 1). Primers that failed any of these validation steps were redesigned and reanalyzed until a suitable primer pair was obtained. Our success rate for primer validation in the first round of design was ~90%.

Primer Design Criteria—The simplicity of the SYBR green qRT-PCR strategy means that effective primer design is critical for selective and efficient template amplification. Because we decided to use gDNA as a validation template during primer design, primers were designed using the largest coding exon as template so that they would effectively amplify both the cDNA target as well as the gDNA validation template. The criteria for primer design were quite restrictive (primer length 20 ± 1 bp, $T_m$ 65 ± 1 °C, and amplimer length 65–75 bp; see “Experimental Procedures”) to allow the use of a single set of amplification conditions for the entire list of genes in a 96-well microtiter plate format qRT-PCR machine.

Methods for cDNA Synthesis—In contrast to standard methods for qRT-PCR that commonly span an intron to reduce the possibility of amplifying contaminating gDNA (55), our use of gDNA as a validation template raised the concern that trace contamination of gDNA in our cDNA samples would contribute to a false-positive qRT-PCR signal. Thus, we performed extensive DNase treatment of our RNA samples prior to cDNA synthesis to eliminate gDNA contamination and confirmed the absence of gDNA in all samples by performing control reactions without reverse transcriptase (No-RT control) before proceeding with high throughput qRT-PCR analyses on the cDNA samples. Because many of the transcripts encoding members of our gene list have been shown to be of low abundance in animal tissues, we optimized our cDNA synthesis...
reactions to maximize template production for qRT-PCR. Comparison of cDNA templates synthesized from either total RNA or poly(A⁺)-mRNA revealed lower Ct values (higher template concentrations) from the mRNA samples (supplemental Fig. 2A). Comparison of cDNA priming methods also revealed that priming with a mixture of oligo(dT) and random hexamers produced lower Ct values (more effective cDNA synthesis) than priming with either oligo(dT) or random hexamers alone (supplemental Fig. 2B), presumably as a result of internal priming events and priming from the transcript poly(A) tail.

Normalization Gene—qRT-PCR transcript abundance data are generally expressed as relative transcript abundance using the ΔΔCt method (41) by comparison to an invariant housekeeping gene that is similarly expressed across all tissues being examined (42). A panel of housekeeping genes was tested for data normalization using cDNAs derived from four adult mouse tissues (supplemental Fig. 3), and ribosomal protein L4 (Rpl4) was selected as the normalization gene because transcript levels were least variant across the four tissues. In addition, triplicate qRT-PCR analyses of Rpl4 were included in all 96-well plate analyses to control for inter-plate and machine variations.

Comparison of qRT-PCR with Microarray Analysis—In an effort to compare data obtained by our qRT-PCR platform with microarray approaches, we analyzed paired RNA samples from wild type C57/BL6 mouse liver and kidney tissues by qRT-PCR and an Affymetrix microarray platform using the GLYCOv2 gene chip from the Consortium for Functional Glycomics (18). As described under “Experimental Procedures,” RMA values from the microarray data yielded normalized intensity values based on the amount of probe hybridized to the array, whereas Affymetrix GCOS data provided present and absent calls for transcripts in each sample. RMA values were compared with the relative transcript abundance data from qRT-PCR analysis (Fig. 1, both plotted on a log₁₀ scale) for 149 glycan-related glycosyltransferase and glycosylhydrolase transcripts called as

FIGURE 2. Relative transcript abundance for genes involved in N-glycan lipid-linked oligosaccharide precursor biosynthesis. A, schematic representation of the N-glycan lipid-linked oligosaccharide biosynthetic pathway was adapted from Ref. 96 (including a key for glycan and lipid structural components in the pathway). Linkages are shown for each step of the biosynthetic pathway, and the numbers in the blue ovals designate the pathway steps in A that link to the transcript abundance data in the corresponding numbered step in B (plotted as a histogram on a log₁₀ scale). Relative transcript abundances for the four mouse tissues are presented as a clustered set of histograms above the corresponding pathway step number (blue numbered oval) and gene names. Multiple genes for a given pathway step are listed in cases where multiple distinct subunits contribute to catalysis or where several genes within a common family encode enzymes capable of creating the specified linkage.

JUNE 20, 2008 • VOLUME 283 • NUMBER 25
JOURNAL OF BIOLOGICAL CHEMISTRY 17303
present (Fig. 1, solid circles) or absent (Fig. 1, open circles) in the microarray analyses. Correlation coefficients between the microarray and qRT-PCR data sets were fairly low ($R^2 = 0.43$ for liver and 0.26 for kidney). Half (50%) of the liver transcripts and about a third (34%) of the kidney transcripts were classified as absent in the microarray analysis (Fig. 1, open circles) consistent with the limited sensitivity of the previously published GLYCOv1 gene chip (14). In contrast, qRT-PCR analysis detected most of the transcripts from both tissue sources (92.6% for liver and 98.7% for kidney) confirming the higher sensitivity of the latter approach. Differences in the dynamic range of the two techniques were also evident, where the microarray data spanned $\sim 2.5$ orders of magnitude as compared with $\sim 7$ orders of magnitude for the qRT-PCR approach.

Integration of Expression Data with Biosynthetic Pathways—In an effort to provide a framework for the interpretation of quantitative data on several hundred glycan-related transcripts, we generated biosynthetic pathway diagrams for the synthesis and modification of all classes of mammalian glycoconjugates and assigned enzymes on the gene list to individual pathway steps. Enzyme assignments were made using the primary literature, several texts in the field (56, 57), and glycan biosynthetic pathway information from the KEGG database (27–29)). Biosynthetic pathway diagrams are shown in Figs. 2–4 and in supplemental Figs. 4–20. Enzyme assignments for the glycan catabolic pathway diagrams are presently underway. For presentation of the qRT-PCR transcript profile data, diagrammatic pathway figures were paired with histogram presentations of relative transcript abundance (on a log10 scale) to allow a visual display of transcripts encoding individual pathway steps. Representative figures are shown for $N$-glycan biosynthesis, including lipid-linked precursor biosynthesis (Fig. 2), trimming and branching reactions (Fig. 3), and complex capping reactions (Fig. 4) for adult mouse kidney, liver, testis, and brain tissues, whereas figures for other glycan biosynthetic pathways can be found in supplemental Figs. 4–20. Additional information on
individual genes for each pathway step can be cross-referenced by gene symbol to the corresponding entries in supplemental Table 1.

For the N-glycan lipid-linked precursor biosynthetic pathway (Fig. 2), we anticipated invariant transcript levels in all four tissues, because this pathway is expected to be constitutive in function across all tissues. Transcript levels were found to vary from gene to gene, but relative expression across the four tissues was similar. Exceptions were genes for which there are more than one isoform (i.e. Alg13, Glt28d2, Stt3a, and Stt3b), where minor differences in tissue-specific expression were observed. Similarly, some processing steps involved in N-glycan trimming and branching are encoded by a single gene isoform (Fig. 3, Mgat1 and Mgat2) and corresponding transcripts were invariant across the four tissues. Other steps in these pathways were accomplished by multiple enzyme isoforms (i.e. Golgi α-mannosidases (Man1a1, Man1a2, Man1c1), B3galT1-6, B4galt1-5), and transcript levels exhibited significant tissue-specific differences for many of the isoforms as previously indicated for these gene families (57–62). Exceptions to this pattern include the single enzyme isoform, Mgat3, adding a bisecting 1,4-GlcNAc, and Fut8, adding a core 1,6-Fuc residue. Transcript levels for both of these enzymes were 10-fold lower in liver compared with kidney, testis, and brain. Considerably greater differences in tissue-specific expression could be identified in the synthesis of the complex capping reactions that are present on the nonreducing termini of N-glycans, O-glycans, and glycolipids (Fig. 4). These capping reactions are generally catalyzed by multiple enzyme isoforms, and the majority of these enzymes exhibit wide differences in tissue-specific expression.

Correlations between qRT-PCR Analysis and N-Linked Glycan Structures—Previous studies correlated glycan structural data...
obtained by MALDI-MS approaches with microarray transcript data from a partial list of glycan-related transcripts (14). In an effort to make a similar comparison with glycan structural data, we compared the transcript abundance from our more extensive glycan-related gene list with the same mouse tissue N-glycan structural data (14). To make a more quantitative comparison between the two data sets, the respective N-glycan MALDI profiles were converted into relative glycan abundance (expressed as a percentage of the total peak intensity) in each tissue as described under “Experimental Procedures” (Fig. 5).

Comparison of relative glycan abundance in the four mouse tissues with the respective transcript profiles (Fig. 6) revealed several correlations (see Table 2 for summary). The previous comparison with microarray transcript profiles demonstrated a correlation between reduced N-glycan core α1,6-Fuc levels in liver (15% of the N-glycan structures in liver compared with 45–52% of the glycan structures in the other tissues) and the inability to detect transcripts encoding Fut8 (14), the enzyme that synthesizes this linkage. Our qRT-PCR data detected Fut8 transcripts in all four tissues, but the greater sensitivity of the latter approach demonstrated an ∼10-fold lower abundance of Fut8 transcripts in liver compared with the other tissues (Fig. 6, A and B). Similarly, prior studies correlated nonreducing terminal α1,3-Fuc residues on N-glycans with the expression of Fut9 (14). Glycan structures containing terminal fucose residues were highly abundant in kidney (37%) and brain (19%) but extremely low abundance in liver and testis (<4%). By comparison our qRT-PCR data indicated abundant Fut9 transcripts in kidney and brain, but no detectable transcripts in liver and testis (Fig. 6, C and D), reflecting a >10^4-fold difference in transcript abundance between the two pairs of tissues.

In addition to these previously identified correlations, we also detected differences in N-glycan structures containing bisecting β1,4-GlcNAc residues in the four mouse tissues. Bisected N-glycan structures included 7% of identified glycans in mouse liver compared with 44, 28, and 29% for kidney, testis, and brain, respectively. This relatively low abundance of bisected structures in liver correlated with >100-fold lower transcript levels for Mbat3 in liver compared with kidney, testis, and brain (Fig. 6, E and F).

Although several correlations could be made between glycan structures and transcript data, there were also cases where the data did not correlate well. For example, oligomannose structures included 85% of the glycans in liver in contrast to 47–55% in the other three mouse tissues (Fig. 6G). Because the predominant oligomannose glycan in liver was a Man5GlcNAc2 structure (59% of oligomannose glycans), one explanation could be greatly reduced expression of N-acetylglucosaminyltransferase 1 (Manat1) in liver. However, transcript abundance for Mbat1 revealed similar levels in liver, kidney, and brain (Fig. 6H). Transcript data for the Golgi α-mannosidases (GMIA (Man1a), GMIB (Man1a2), and GMIC (Man1c1)), which produce the Man5GlcNAc2 structure, were also not significantly reduced in liver compared with the other tissues (Fig. 6I). A likely explanation for elevated oligomannose structures in liver is the extensive elaboration of the smooth endoplasmic reticulum in liver hepatocytes to provide the enzymatic machinery for drug detoxification and lipid biosynthesis (63, 64). This highly abundant, specialized endomembrane system is enriched in glycosylated enzymes that will not likely encounter the Golgi N-glycan processing machinery required for conversion of oligomannose structures into highly branched complex N-glycans. Thus, glycan abundance in this tissue likely reflects the unique organelle structure in hepatocytes rather than the glycan processing machinery in the Golgi complex of these cells.

The glycan structural data also indicated a considerably reduced level of sialylation in the kidney compared with the other tissues (14). In contrast, transcripts encoding all of the sialyltransferase genes (GT29 CAZy family) are detected at some level in kidney, with the exception of St6Gal2, which was highly expressed in brain and barely detectable in kidney (Fig.
However, four members of the GT29 family are proposed to be the major contributors to N-glycan H2,3- or H2,6-sialylation (St3gal3, St3gal4, St3gal6, and St6gal1) (14). Of these four enzymes, only St3gal4, involved in the H2,3-sialylation of type II (Galβ1–4GlcNAc) sequences, is slightly reduced (~5-fold) in kidney (Fig. 7). These data suggest that decreased transcript levels for the sialyltransferases in kidney are not likely to be the major contributor to reduced sialylation in this tissue. Additional potential contributors to reduced sialylation in kidney include post-transcriptional control of sialyltransferase activities, increased removal of sialic acid residues by neuraminidases, or decreased CMP-sialic acid precursor pools. Transcript abundance for the four neuraminidase/sialidase genes (GH33 family, Fig. 7) indicates increased levels for the lysosomal neuraminidase, Neu1 (~6-fold higher in kidney), and the cytosolic neuraminidase, Neu2 (~8-fold higher in kidney). Transcript abundance for the membrane sialidases, Neu3 and Neu4, revealed lower (11-fold) or undetectable levels in kidney, respectively. Transcript abundance for enzymes involved in sugar nucleotide biosynthesis leading to CMP-sialic acid precursor production (Renbp, Nans, and Cmas, supplemental Fig. 4) and CMP-sialic acid transport in the Golgi complex (Slc35a1, supplemental Fig. 19) were either higher in kidney or similar in all four mouse tissues, suggesting that depleted precursor pools in kidney were unlikely to account for reduced sialylation in this tissue.

Correlations of Transcript Profiles with Low Abundance Glycan Structures—Tissue-specific expression of enzymes that synthesize low abundance glycan structures can also be detected among the qRT-PCR transcript profiles. These data can be compared with glycan structures previously published in the literature as well as previously published transcript abundance data or tissue-specific EST abundance (available through the UniGene EST ProfileViewer). Examples of similarities between our qRT-PCR data and prior published data include Northern blots detecting highly abundant transcripts for Mgat3 in kidney and brain in contrast to moderate levels in testis and barely detectable transcripts in liver (65). This pattern of transcript abundance is qualitatively similar to our qRT-PCR results (Fig. 6), but the qRT-PCR data provide quantitative val-


## Table 2: Summary of correlations between transcript profiles and glycan structures in mouse tissues

| Glycan | Gene(s) | Expression pattern | Correlation | Comments |
|--------|---------|--------------------|-------------|----------|
| 1,3-Fuc transferases | Fut8 | Low (5-fold lower) in liver compared with kidney | Positive | N/H9251 | 1,6-Fuc, extremely low (4%) in liver and testis, consistent with loss of SSEA-1 staining in liver (67) |
| Oligomannose glycan structures | Mgat3 | Lower abundance in liver | Positive | Bisecting | Reduced N/H9251, oligomannose glycans correlate with complex N-glycans |
| Oligomannose glycan structures | Mgat1, Man1a1 | High abundance of oligomannose glycans | None | Ratio of high N/H9251- and high N/H9252-mannosidase levels in liver compared with kidney, testis, and brain |
| Trimannosyl core | Mgat1, Man1c1 | Reduced abundance of mannose | None | Reductive | Combination of reduced N/H9251 and elevated N/H9252 may account for the absence of sialylated glycans in liver |
| Sialylated N-glycans | Neu1–4, St3gal4 | Reduced abundance of sialylated structures in kidney, other sialyltransferases unchanged or increased | Inverse or complex | Sialylation may regulate homotypic NCAM adhesion |

Transcript Profiling of Mouse Glycan-related Genes

The murine gene responsible for A/B histo-blood group antigen biosynthesis is a cis-AB glycosyltransferase (Ab0) that can synthesize both α1,3GalNAc (blood group A) and α1,3Gal (blood group B) linkages (74). Humans synthesize A/B antigens broadly in the epithelial cells of gastrointestinal, esophageal, bronchopulmonary, oral, and urogenital tissues and bone marrow progenitors in contrast to low expression in liver, spleen, and kidney and no detectable expression in brain and muscle (57). In contrast, mice appear to express Ab0 in the urogenital tract and intestine based on our qRT-PCR data (Fig. 4, high in testis, base-line levels in kidney, liver, and brain) and EST abundance data (only detected in bladder, epididymus, prostate, testis, and intestine). These transcript data are consistent with the absence of detectable A/B antigens in murine blood cells and salivary secretions based on agglutination tests and immunologic detection methods (74).
Transcript Profiling of Mouse Glycan-related Genes

FIGURE 7. Relative transcript abundance for sialyltransferases and neuraminidases in mouse tissues. Sialyltransferases are grouped according to the type of linkage they create upon transfer from a CMP-sialic acid donor and respective gene names for the sialyltransferases and neuraminidases are listed at the bottom of the histogram. Relative transcript abundance data are plotted as a clustered set of histograms for the respective genes as indicated for kidney (black bars), liver (dark gray bars), testis (light gray bars), and brain (white bars). Error bars represent the mean ± 1 S.D.

bers (B3gnt1–8) and β1,4Gal transferases family members (B4galt1–7) to generate i-blood group antigens. Substrate specificities for recombinant B3gnt1–8 have been characterized (75), but their roles in polylactosamine extension in vivo are still unclear (76). In adult mouse tissues all of the B3gnt family members are expressed widely (Fig. 4). The B3gnt2 isofrom is particularly highly expressed in all four mouse tissues examined. This latter enzyme has strong polylactosamine synthesizing activity (75), and recent studies on B3gnt2-deficient mice indicate a marked reduction in polylactosamine on N-glycans in immunological tissues and induction of a hyper-reactive immune response (76).

Branching of the linear polylactosamine chains is accomplished by a collection of β1,6GlcNAc transferases (Gcnt1–3) to generate I-antigen structures (57). I-GlcNAc transferase (IGNt) activity is largely accomplished by the broadly expressed Gcnt2, which can be found as three splice variants (77). Gcnt2 was highly expressed in all four mouse tissues examined (Fig. 4) consistent with the loss of IGNt activity and I antigen structures in Gcnt2-deficient mice (77). Core2GlcNAcT-II (Gcnt1) predominantly forms 2 branches in O-glycans but has weak IGNt activity (78). Our data suggest that Gcnt1 is widely expressed, but its contribution to polylactosamine branching remains to be resolved. A third isofrom, Gcnt3, has been isolated recently and shown to have core 2, core 4, and IGNt activity (79). Transcripts encoding this latter enzyme are highly expressed in testis, moderately expressed in kidney, and exceptionally low in abundance in liver and brain (Fig. 4).

The synthesis of human Sda/CT/Cad antigens is accomplished by the transfer of a β1,4GalNAc residue to a disaccharide acceptor to form the terminal GalNAcβ1,4[ NeuSAcα2,3]Galβ trisaccharide on red blood cells, body fluids, kidney, and large intestine (80). The enzyme that generates this linkage, B4galnt2, is onco-developmentally regulated, increasing activity with age, and down-regulated in colon carcinoma (81). In mice, B4galnt2 is poorly expressed in kidney, liver, and brain, whereas moderate expression is found in testis based on our qRT-PCR data (Fig. 4), and EST abundance data suggest additional elevated expression in mouse intestine.

Sulfation of N-glycan and mucin-type O-glycan structures is restricted to GalNAc-4-SO4 addition to terminal structures in glycoprotein hormone N-glycans, HNK-1 epitope antigen biosynthesis, sulfation of polylactosamine structures to form keratan sulfate, and sulfation of terminal GlcNAc residues on sialyl 6-sulfo Lewis x structures (510-fold lower transcript levels in testis compared to polylactosamine branching remains to be resolved. A third

FIGURE 7. Relative transcript abundance for sialyltransferases and neuraminidases in mouse tissues. Sialyltransferases are grouped according to the type of linkage they create upon transfer from a CMP-sialic acid donor and respective gene names for the sialyltransferases and neuraminidases are listed at the bottom of the histogram. Relative transcript abundance data are plotted as a clustered set of histograms for the respective genes as indicated for kidney (black bars), liver (dark gray bars), testis (light gray bars), and brain (white bars). Error bars represent the mean ± 1 S.D.

Branching of the linear polylactosamine chains is accomplished by a collection of β1,6GlcNAc transferases (Gcnt1–3) to generate I-antigen structures (57). I-GlcNAc transferase (IGNt) activity is largely accomplished by the broadly expressed Gcnt2, which can be found as three splice variants (77). Gcnt2 was highly expressed in all four mouse tissues examined (Fig. 4) consistent with the loss of IGNt activity and I antigen structures in Gcnt2-deficient mice (77). Core2GlcNAcT-II (Gcnt1) predominantly forms 2 branches in O-glycans but has weak IGNt activity (78). Our data suggest that Gcnt1 is widely expressed, but its contribution to polylactosamine branching remains to be resolved. A third isofrom, Gcnt3, has been isolated recently and shown to have core 2, core 4, and IGNt activity (79). Transcripts encoding this latter enzyme are highly expressed in testis, moderately expressed in kidney, and exceptionally low in abundance in liver and brain (Fig. 4).

The synthesis of human Sda/CT/Cad antigens is accomplished by the transfer of a β1,4GalNAc residue to a disaccharide acceptor to form the terminal GalNAcβ1,4[ NeuSAcα2,3]Galβ trisaccharide on red blood cells, body fluids, kidney, and large intestine (80). The enzyme that generates this linkage, B4galnt2, is onco-developmentally regulated, increasing activity with age, and down-regulated in colon carcinoma (81). In mice, B4galnt2 is poorly expressed
with high endothelial venules, where they create ligands for interaction with L-selectin during lymphocyte homing to lymph nodes (91, 92). The sulfated structures are generated on the termini of core 1 and core 2 sialomucins by the sulfotransferases GlcNAc6ST-1 (Chst2) and GlcNAc6ST-2 (Chst4) (92). Expression of Chst4 was found to be restricted in expression pattern to high endothelial venules and a limited number of other sites by Northern blotting (93) consistent with low expression levels for transcripts in the four mouse tissues by qRT-PCR (Fig. 4). In contrast, Chst2 is more broadly expressed in mouse tissues based on Northern blots (94) consistent with high transcript levels in brain and moderate levels in kidney, liver, and testis by qRT-PCR (Fig. 4).

Additional transcript abundance data were collected for other glycan classes (see supplemental figures), and in instances where biochemical pathways could be mapped, the data are presented with the corresponding pathway diagrams. In cases where pathway data are not relevant (lectins, sugar/sugar-nucleotide transporters, and proteoglycan core proteins), the data are presented in histogram form based on the classification of the proteins.

**DISCUSSION**

Whereas protein-carbohydrate interactions are known to play critical roles in biological recognition events, very little is known about the global regulation of glycan biosynthesis and catabolism. The complex interplay between glycosyltransferase expression, sugar precursor availability, competition among enzymes for acceptor modification, expression levels of acceptor glycoproteins and glycolipids, and rates of glycoprotein catabolism all contribute to glycan composition of a given glycoconjugate on cell surface or secreted macromolecules. Our goal in this study was to examine correlations of glycan structural data with detailed transcript profiles of enzymes and proteins involved in glycan synthesis, modification, recognition, and catabolism in an effort to test the hypothesis that glycan structures are predominantly regulated at the transcriptional level. To accomplish this task, two critical components were required to make the comparisons: sensitive methods for broad-based glycan structural analysis paired with sensitive and quantitative methods for profiling the full range of transcripts encoding the machinery for glycan biosynthesis, recognition, and catabolism. The goal of broad-based glycan profiling has benefited from recent advances in glycan structural analysis (15–17). In contrast, standard high throughput methods for transcript analysis (i.e. microarray and SAGE approaches) are restricted in sensitivity and dynamic range necessary to quantitate the low abundance transcripts that are commonly involved in glycan biosynthesis and modification (14, 18–22, 24). We have developed an alternative approach employing a medium-throughput qRT-PCR platform for the analysis of >700 glyobiology-related genes that has a quantitative dynamic range of ~7 orders of magnitude and the ability to detect very low abundance transcripts.

Direct comparison of our qRT-PCR platform with microarray analyses using the Affymetrix GLYCOv2 microarray approach (18) revealed an improved detection of low abundance transcripts (2 times more transcripts detected in liver and 1.5 times more detected in kidney). Among those transcripts that were detected in the microarray analyses, the general trends for the microarray and the qRT-PCR analyses were in the same direction, but the correlation coefficients were quite poor ($R^2 = 0.39$ for liver and 0.24 for kidney), confirming the common use of qRT-PCR to provide quantitative validation of transcript changes initially detected using microarray methods (53). The Consortium for Functional Glycomics has now generated a new version of their GLYCOv3 gene chip, and comparisons of paired analyses by qRT-PCR and the new chip may yield better correlations in the future.

Several lines of evidence strongly argue that our qRT-PCR data reflect the true quantitative measure of transcript abundance in the respective tissues. First, our qRT-PCR amplifications have a linear response in a dilution series of both genomic DNA templates and cDNA templates across the entire dynamic range of analysis. Second, our PCR efficiencies have been confirmed both through a template dilution series as well as by determining the rate of amplimer appearance using the LinRegPCR program (43). In any instance where an amplification efficiency value deviates from our restrictive criteria ($100 \pm 10\%$) during either primer validation or sample analysis, the experiments are either repeated or subjected to primer redesign until our efficiency criteria are obtained. Our restrictive primer design and validation criteria have allowed the use of uniform amplification conditions for analysis of the entire gene list and yielded validated primer sets for ~90% of the target genes in the first round of design. Most of the primer sets that failed validation in the first round of design could be effectively validated upon redesign. Finally, in each instance where we have compared our qRT-PCR data with literature data on transcript abundance based on Northern blots, the published data agree with our qRT-PCR results, except where the sensitivity of the Northern blots is too low for transcript detection.

The key components in our transcript analytical platform were the collation of a comprehensive target list for glycobiology-related genes, the development of protocols and validated primer sets for the qRT-PCR approach, and the assignment of members of the gene list to discrete steps in their respective biosynthetic pathways. Collation of the comprehensive gene list was more difficult than anticipated because of the complexity of glycan biosynthetic and modification pathways. Numerous steps in glycan biosynthetic pathways are catalyzed by multiple enzyme isoforms, many of which are not well characterized in regard to acceptor or donor specificity. As a result it is difficult to achieve one-to-one mapping of enzymes to discrete pathway steps without detailed knowledge of the primary literature and the specificity of a given enzyme system. Although web resources such as the CAZy and KEGG databases provided effective starting points for gene list assembly or integrated pathway mapping, there is not a single, publicly available resource for all of the relevant target genes in a form that contains functional annotations or effective pathway mapping. Thus, extensive manual annotation of the gene list and pathway diagrams was required to create a bridge between our glycan-related transcript abundance data and corresponding glycan structural data. We consider the annotated pathway diagrams and the underlying gene list to be a “work in progress,”
because new enzymes are constantly being added to the list and ongoing characterization of the existing members of the list will likely lead to revisions of the enzyme assignments in the pathway diagrams. However, the present gene list is still more than double the size of the largest glycan-related gene list used to create focused microarrays (i.e. Consortium for Functional Glycomics GLYCOv1 gene chip) employed in prior transcript analysis studies (14). We have created a web site for archiving our glycotranscriptome data (NIH-NCRR Integrated Technology Resource for Biomedical Glycomics) containing a catalog of the updated gene lists, primer design information, pathway diagrams, and eventual archiving of transcript data sets generated by our qRT-PCR analysis.

Application of the qRT-PCR methodology to mouse tissue transcript analysis revealed a broad diversity of tissue-specific transcript expression for glycan-related enzymes and proteins. Several pathways (i.e. N-glycan lipid-linked precursor synthesis (Fig. 2) and GPI anchor biosynthesis (supplemental Fig. 7)) are catalyzed predominantly by single enzyme isoforms and generally exhibited very little tissue-specific variation across the linear pathway steps. Other pathways, especially those that elab-
orate terminal glycan epitopes on N-glycans, O-glycans, glycolipids (Fig. 4), and sulfated proteoglycans on cell surface glycoconjugates (supplemental Figs. 11–13), were catalyzed by multiple enzyme isoforms that exhibited complex tissue-specific expression patterns. This proliferation of isoforms with varying specificities and expression patterns likely reflects the varying diversity in functions of cell surface glycoconjugates in neural tissues (Fig. 7 and supplemental Fig. 12). Widely variant tissue-specific expression patterns can be seen in the biosynthesis of various enzyme isoforms. Some enzymes involved in the sulfation of heparan sulfate exhibit little variation across the four tissues (Ndst1, Ndst2, Hs2st1, Hs6st1, Hs3st3a1, and Hs3st6, supplemental Fig. 11), whereas other isoforms exhibit profound tissue-specific expression. Similar combinations of invariant and tissue-specific isoforms can be seen in the biosynthesis of chondroitin, dermatan, and keratan sulfates (supplemental Figs. 12 and 13). Widely variant tissue-specific expression patterns can also be found for the core proteins that act as acceptors for heparan, chondroitin, dermatan, and keratan sulfates (supplemental Fig. 14). In a more focused study, we previously demonstrated that proteoglycan sulfation and core protein levels vary during mouse embryonic stem cell differentiation, whereas the proteoglycan core and disaccharide co-polymer extension enzymes are not altered in expression during differentiation (95). Thus, glycan-related transcript profiling can provide insights into which enzyme isoforms are regulated in a given biological system.

Transcript profiling data can also highlight points of transcriptional regulation that may be unexpected, including the expression of St6sia2 and NCAM in testis in addition to its expected location in neural tissues (Fig. 7 and supplemental Fig. 20). The appearance of nonpolysialylated NCAM in testis has been reported (70, 71), but the presence of both St6sia2 and NCAM transcripts in testis suggests an additional role for polysi-
ylation in this latter tissue. Similarly, expression of the ganglioside α2,8-sialyltransferases, St6sia2 and St6sia3, in testis is consistent with recent studies demonstrating a critical role for complex gangliosides in mouse spermatogenesis (73).

Although our initial profiles of glycan-related transcripts in four mouse tissues provide a focused view of glycan regulation, one of our eventual goals is to generate a more extensive map of glycan-related transcript expression in a broad collection of mouse tissues to correlate these data with parallel sets of quan-
titative glycan structural data. This atlas of glycan and tran-
transcript expression will act as a framework for understanding regulation of glycan structures in animal systems. Further developments in glycan analytical methodologies will be required to attain these goals, particularly for O-glycans, glycolipids, and proteoglycans, where chemical and enzymatic release methods are less developed for high throughput quantitative glycan profiling. Additional work must also be done to complete the enzyme assignments for glycan-related pathways, particularly in the catabolism of all glycan classes. Steady state levels of some glycoproteins will likely be regulated by a balance of glycan synthesis and catabolism. Thus, modeling of glycan abundance in biological systems will require detailed knowledge of the mechanisms for regulating both biosynthetic and catabolic pathways.

In conclusion, we have initiated studies to test the hypothesis that glycan expression in animal tissues is regulated predominantly at the transcriptional level. To accomplish this we have developed a robust, flexible platform for transcript analysis of mouse glycan-related enzymes and proteins with a wide quantitative dynamic range. Insights that come from the ability to quantify the full range of glycan-related transcripts revealed numerous correlations with glycan structural data indicating that many, but not all, glycan structural changes can be accounted for by transcriptional regulation of the glycan synthetic machinery. Additional use of this analytical platform in other biological systems should complement glycan structural studies and lead to a greater understanding of the roles that glycans play in animal physiology and pathology.

Acknowledgments—We thank Drs. Mary Bedell and Nancy Manley (Department of Genetics, University of Georgia) for supplying mouse tissues and genomic DNA and the "army" of undergraduates who helped design and validate primers: Roy Johnson, Christina Seibert, Jayne Littlejohn, Ken Chen, Karla Flores, April Oxford, and Michael Humphreys. Special thanks to Dr. Paristoo Azadi and Dr. Stuart Humphreys. We also thank the Consortium for Functional Glycomics Core E Director Steven Head and Tim Gilmarty for correspondence regarding the organization of the lectin families. We thank Dr. Kurt Drickamer and the Consortium for Functional Glycomics supported by National Institutes of Health NIGMS Grant U54 GM62116.

REFERENCES

1. Varki, A. (1993) Glycobiology 3, 97–130
2. Lowe, J. B. (2001) Cell 104, 809–812
3. Lowe, J. B., and Marth, J. D. (2003) Annu. Rev. Biochem. 72, 643–691
4. Ohtsubo, K., and Marth, J. D. (2006) Cell 126, 855–867
5. Schachter, H. (2000) Glycoconj. J. 17, 465–483
6. Brockhausen, I., Schutzbach, J., and Kuhns, W. (1998) Acta Anat. 161, 36–78
7. Dennis, J. W., Granovsky, M., and Warren, C. E. (1999) BioEssays 21, 412–421
8. Haltiwanger, R. S., and Lowe, J. B. (2004) Annu. Rev. Biochem. 73, 491–537
9. Taniguchi, N., Miyoshi, E., Gu, J., Honke, K., and Matsumoto, A. (2006) Curr. Opin. Struct. Biol. 16, 561–566
10. Kornfeld, R., and Kornfeld, S. (1985) Annu. Rev. Biochem. 54, 631–664
11. Schachter, H. (1984) Clin. Biochem. 17, 3–14
12. Schachter, H. (1984) Biol. Cell 51, 133–145
13. Schachter, H. (1991) Glycobiology 1, 453–461
14. Comelli, E. M., Head, S. R., Gilmarty, T., Whisentant, T., Haslam, S. M., North, S. J., Wong, N.-K., Kudo, T., Narimatsu, H., Esko, J. D., Drickamer, K., Dell, A., and Paulson, J. C. (2006) Glycobiology 16, 117–131
15. Aoki, K., Perlman, M., Lim, J. M., Cantu, R., Wells, L., and Tiemeyer, M. (2007) J. Biol. Chem. 282, 9127–9142
16. Kui Wong, N., Easton, R. L., Panico, M., Sutton-Smith, M., Morrison, J. C., Lattanzio, F. A., Morris, H. R., Clark, G. F., Dell, A., and Patankar, M. S. (2003) J. Biol. Chem. 278, 28619–28634
17. Sutton-Smith, M., Morris, H. R., Grewal, P. K., Hewitt, J. E., Bittner, R. E., Goldfin, E., Schifflmann, R., and Dell, A. (2002) Biochem. Soc. Symp. 69, 105–115
18. Bax, M., Garcia-Vallejo, J. J., Jang-Lee, J. N., North, S. J., Gilmarty, T. J., Hernandez, G., Crocker, P. R., Leffler, H., Head, S. R., Haslam, S. M., Dell, A., and van Kooiy, Y. (2007) J. Immunol. 179, 8216–8224
19. Hemmoranta, H., Satomaa, T., Blomqvist, M., Heiskanen, A., Aitio, O., Saarinen, J., Natunen, J., Partanen, L., Jäntti, L., and Jasim, T. (2007) Exp. Hematol. 35, 1279–1292
20. Naito, Y., Takematsu, H., Koyama, S., Miyake, Y., Yamamoto, H., Fujinawa, R., Sugai, M., Okuno, Y., Tsujimoto, G., Yamaji, T., Hashimoto, Y., Itohara, S., Kawasaki, T., Suzuki, A., and Kozutsumi, Y. (2007) Mol. Cell. Biol. 27, 3008–3022
21. Smith, F. I., Qu, Q., Hong, S. I., Kim, K. S., Gilmarty, T. J., and Head, S. R. (2005) Gene Express. Patterns 5, 740–749
22. Yamamoto, H., Takematsu, H., Fujinawa, R., Naito, Y., Okuno, Y., Tsujimoto, G., Suzuki, A., and Kozutsumi, Y. (2007) Plos ONE 2, e1232
23. Young, W. W., Jr., Holcomb, D. R., Ten Hagen, K. G., and Tabak, L. A. (2003) Glycobiology 13, 549–557
24. Ishii, A., Ikeda, T., Hitoshi, S., Fujimoto, I., Torii, T., Sakuma, K., Nakakita, S., Hase, S., and Ikenaka, K. (2007) Glycobiology 17, 261–276
25. Coutinho, P. M., Deleury, E., Davies, G. J., and Henriissat, B. (2003) J. Mol. Biol. 328, 307–317
26. Coutinho, P. M., and Henriissat, B. (1999) in Recent Advances in Carbohydrate Bioengineering (Gilbert, H. I., Davies, G., Henriissat, B., and Svensson, B., eds) pp. 3–12, Royal Society of Chemistry, Cambridge, UK
27. Kanekhasa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Ioh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. (2008) Nucleic Acids Res. 36, D480–D484
28. Kanekhasa, M., and Goto, S. (2000) Nucleic Acids Res. 28, 27–30
29. Kanekhasa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006) Nucleic Acids Res. 34, D354–D357
30. Diehn, M., Sherlock, G., Binkley, G., Jin, H., Mateas, J. C., Hernandez-Boussard, T., Rees, A., Cherry, J. M., Botstein, D., Brown, P. O., and Alizadeh, A. A. (2003) Nucleic Acids Res. 31, 219–223
31. Garcia-Vallejo, J. J., Gringhuis, S. I., van Dijk, W., and van Die, I. (2006) Methods Mol. Biol. 347, 187–209
32. Garcia-Vallejo, J. J., Van Dijk, W., Van Het Hof, B., Van Die, I., Engelse, M. A., Van Hinsbergh, V. W., and Gringhuis, S. I. (2006) J. Cell. Physiol. 206, 203–210
33. Taylor, M. E., and Drickamer, K. (2006) Introduction to Glycobiology, 2nd Ed., pp. 122–175, Oxford University Press, Oxford
34. Saier, M. H., Jr., Tran, C. V., and Barabote, R. D. (2006) Nucleic Acids Res. 34, D181–D186
35. Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Carnes, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., Ger, L. Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Ostell, J., Miller, V., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R. L., Tatusov, T. A., Wagner, L., and Yasinchenko, E. (2007) Nucleic Acids Res. 35, D5–D12
36. Maglott, D., Ostell, J., Pruitt, K. D., and Tatusov, T. (2007) Nucleic Acids Res. 35, D26–D31
37. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410
38. Rozen, S., and Skaltsky, H. (2000) in Bioinformatics Methods and Protocols: Methods in Molecular Biology (Krawetz, S., and Misener, S., eds) pp.
