The Caenorhabditis elegans Gene, gly-2, Can Rescue the N-Acetylglucosaminylntransferase V Mutation of Lec4 Cells*

Received for publication, February 11, 2002, and in revised form, March 28, 2002
Published, JBC Papers in Press, April 5, 2002, DOI 10.1074/jbc.M201390200

Charles E. Warren‡, Aldis Križus‡, Peter J. Roy‡¶§, Joseph G. Culotti¶§, and James W. Dennis¶§

From the ‡Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5 and ¶Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, M5S 1A8, Canada

UDP-N-acetylglucosamine:α-6-mannoside β-1,6-N-acetylglucosaminylntransferase V (GlcNAc-TV) is one of a set of sequence-unrelated GlcNAc-T enzymes that create branches in complex-type N-glycans (1). These branches can be further elongated by galactosyltransferase and other enzymes to create the mature glycoprotein oligosaccharides. The GlcNAcβ1,6 branch resulting from GlcNAc-TV action is distinct in that it is the preferred site for elongation with polylactosamine chains, repeating lactosamine units that themselves can be further branched and carry a variety of terminal structures. GlcNAc-TV is thus a potential regulator of polylactosamine transfers in that it has a specific temporal and spatial expression pattern in the developing mouse embryo. Expression is concentrated in neuronal tissues, specialized epithelium, and regions with stem cell-like populations. Zygotropic expression increases at about 9.5 days post coitus, which coincides with the onset of organogenesis (2).

Mice deficient in GlcNAc-TV activity through mutation of the Mgt-5 locus are viable but develop glomerulonephritis with age, which is associated with T cell hypersensitivity, apparently as a result of altered activation kinetics of the T cell receptor complex (3). When the Mgt-5° allele is combined with a mouse mammary tumor virus-promoted Polyomavirus middle T antigen transgene, multifocal tumorigenesis is delayed, and metastasis caused by the Polyomavirus middle T antigen is dramatically suppressed (4). This result is consistent with prior observations that tumor cell lines selected by resistance to the cytotoxic lectin Phaseolus vulgaris leucoagglutinin (L-PHA) deficient in GlcNAc-TV also failed to metastasize in syngeneic mice (5).

Although the Mgt-5° mouse is highly informative, systematic analysis of a complex viable phenotype remains difficult, particularly the identification of the dependent molecules and pathways. We therefore sought a simpler model organism in which synthetic genetics could be carried out rapidly to characterize the complex pleiotropic phenotypes expected from disruption of the glycosylation machinery. Because of the cellular non-autonomy typical of glycosylation phenotypes and of the phylogenetic restriction of complex-type N-glycans to metazoans, a whole animal model is necessary. Caenorhabditis elegans is the simplest and most highly characterized animal, its adult anatomy and developmental lineage have been completely determined (6, 7), and its genome is essentially completely sequenced (8). C. elegans is highly tractable to experimental phenotypic and genetic analysis, and there are numerous examples demonstrating that genetic pathways found in mammals are also conserved in this nematode (9–13).
Surveys of the *C. elegans* genome sequence revealed a coding potential for most known glycosyltransferase genes (14). Genes encoding active polypeptide GalNAc-transferases (15), GlcNAc-TI (16), and a fucosyltransferase (17) have been characterized. In addition, there are at least three *sog* genes that are elements of a proteoglycan glycosylation pathway that when mutated cause severe and pleiotropic defects (18–21). A recent NMR-mass spectrometry study identified the abundant N- and O-glycans in *C. elegans* (22). The canonical oligomannose series of N-glycans were observed, but atypical O-glycans were found where polypeptide linked GalNAc was β1–6-branched as in mammals but substituted with glucose or galactose rather than GlcNAc. We characterized the 6 homologues of core 2 GlcNAc-T (23) and demonstrated that gly-1 transfers glucose from UDP-glucose to core 1 acceptor consistent with the inference based on the structural analysis (24).

We observed that the *C. elegans* genome encodes a single gene, designated *gly-2*, which is homologous to mammalian GlcNAc-TV sequences. In this paper, we establish that the nematode orthologue is functionally equivalent to that from *mammals and that* *C. elegans* is an appropriate model in which to pursue investigations of the contributions to fitness by β6-GlcNAc-T branched glycans.

**EXPERIMENTAL PROCEDURES**

**Strains and Materials—** Primer sequences (ACGT Corp.) are available on request. The wild-type Bristol strain of *C. elegans* (N2) (25) and *him-5(e1490)* were available as laboratory stocks. CB1282 *dpv-5(e1292) IV, DR466 him-5(e1490) V*, and DR435 *dpv-5(e611) unc-13(e511) I* were supplied by T. Sternagel (*C. elegans* Genetics Center, University of Minnesota). BC107 btl-4(e937), dpv-14(e1881) I was the gift of Dr. D. Baillie (Simon Fraser University, Vancouver, Canada). Standard husbandry methods were used (27, 26). Cosmid C55B7 was obtained from Dr. A. Coulson (Sanger Centre, Cambridge, UK). pPD95.77 and pPD95.89 were from Dr. A. Fire (Carnegie Institute of Washington). pMH#6 was obtained from Dr. M. Han (University of Colorado). pRF4 originated in the laboratory of Dr. J. Kramer (Northwestern University Medical School), and pMKRF1 was the gift of Dr. F. Hagen (University of Rochester Medical School). pCMVCD20 was obtained from Dr. E. Harlow (Massachusetts General Hospital Cancer Center), and Dr. P. Stamatoyannopoulos (University of Washington). Primer sequences (ACGT Corp.) are available on request. 5’ RACE—Poly(A)^+ RNA was isolated from mixed populations of *C. elegans* using a QuickPrep Micro mRNA purification kit (Pharmacia). The 5’ RACE system (Invitrogen) was used according to the manufacturer’s instructions. First strand cDNA synthesis was primed with yk5’rc1. First round PCR using AmpliTaq Gold (PerkinElmer Life Sciences) was primed with yk5’rc1. The second round PCR used Pfu DNA polymerase (CLONTECH) and yk5’rc2 as the gene-specific primer. Amplimer was sequenced directly and subcloned into the EcoRV site of pBZRv-O-2 (Invitrogen). Independent recombinants were analyzed by colony PCR using SL1, SL2, or RACE anchor and yk5’rc2 primers.

**Northern Analysis—** Non-starved mixed stage animals from Bristol N2 and *him-5(e1490)* strains were used to prepare poly(A)^+ RNA using a Dynabeads kit (Dynal A. S.) after disruption in a Polytron (Kinetica). ~1 μg of RNA was fractionated, blotted, probed with the α-28-labeled *Sal*I/*Smal*I fragment of yk126h8, and analyzed with a PhosphorImager (Storm/ImageQuant, Molecular Dynamics).

**Construction of Mammalian Expression Vectors—** pISTH1 was constructed from pMKRF1 (15) by replacing the *Not*I-*Bam*HI segment upstream of the cloning site with an *Nde*I-*Bgl*II fragment from pCITE4b (+) (Novagen). N-terminal truncations of GLY-2 were generated by PCR from yk126h8 as template using *Pfu* DNA polymerase primed by yk5’rc3 and one of yk13L-f, gly-2-1332, gly-2-1317, or gly-2-1338. Amplimers were subcloned into the EcoRV site of pBZRv-O-2 (Invitrogen) and sequenced. *Bam*HI-fragments of error-free subclones were ligated into the *Bam*HI site of plSTH1. Ligation junctions, frame, and orientation were confirmed by DNA sequencing. A yk5’rc2 and SL1-primer TaqDNA polymerase PCR product of the RACE amplimer was subcloned into EcoRV cut and Tailed pGEM3zf (+) (Promega) forming pYS. pCDNA3: yk126h8 was created by subcloning the *Pvu*I-*Smal*I fragment of pCDNA3: *Eco*RI-PCR product into pCSYK-1. A truncated expression construct for mature SL1 trans-spliced CDNA (pCSYK-1) was constructed by combining the *Spe*I-*Not*I fragment of pYS with the *Not*I-*Not*I fragment of pCDNA3: yk126h8 (+) in *Spe*I-*Not*I cut pZEr0-2, the *Bam*HI-*Not*I fragment of which was subcloned into pCDNA3. The amplified region and ligion junctions were checked by DNA sequencing. The *Not*I-PCR product of yk1110R (mutated) was introduced into pCSYK-1 by homologous recombination directed by primer GLY2-L116R using the Chameleon kit according to the manufacturer’s instructions (Stratagene). The complete transcriptional unit of the resulting construct, pCSYK-L116R, was sequenced. pEGFP-GLY2 was subcloned by subcloning the ykR_r product generated by PCR amplification with *Pfu* DNA polymerase from yk126h8 template into the *Bam*HI site of pEGFP-C3 (CLONTECH). The insert and ligion junctions were completely sequenced and found to be in-frame and error-free. pLNC5-L116R was derived by replacing the *Bsr*XI-EcoRV fragment with the equivalent section of pCSYK-L116R to generate pEGFP-L116R. The inserted segment was confirmed by sequencing.

**Transient Expression and Secretion of GLY-2 in Lec2 Cells—** Lec2 cells (Invitrogen) in 1 ml of OptiMEM-I (Invitrogen)/well. One ml of α-minimal essential medium containing 20% fetal bovine serum was added to the wells, and the clusters were transferred to a humidified 5% CO2 atmosphere at 30 °C overnight. The following day, wells contents were aspirated and replaced with 2 ml of α-minimal essential medium containing 10% fetal bovine serum, and incubation was continued until 78 h post-transfection. Conditioned medium was clarified by centrifugation at 1800 × g for 10 min and stored at 4 °C after the addition of sodium azide to 0.05% v/v.

**Immunopurification of Recombinant Proteins—** Recombinant proteins were directed by pSTH1-based plasmids bear an N-terminal S-tag that was assayed according to the manufacturer’s instructions in conditioned media from the transient transfections (Novagen). 1.25 pmol of recombinant protein in conditioned medium was immunoprecipitated and diluted into 1 ml of dilution buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.025% w/v LipofectAMINE (Invitrogen) in 1 ml of OptiMEM-I (Invitrogen)/well. One ml of α-minimal essential medium containing 20% fetal bovine serum was added to the wells, and the clusters were transferred to a humidified 5% CO2 atmosphere at 30 °C overnight. The following day, wells contents were aspirated and replaced with 2 ml of α-minimal essential medium containing 10% fetal bovine serum, and incubation was continued until 78 h post-transfection. Conditioned medium was clarified by centrifugation at 1800 × g for 10 min and stored at 4 °C after the addition of sodium azide to 0.05% v/v.

**Glycan Analysis of Lec4 Cells Transfected with gly-2—** Enzyme activity was measured using synthetic specific acceptors (28). Assays contained 1 ml *βGlcNAc1,2*-*Man*(1,6)*βGlc-O(CH2)7CH3*, CH3 acceptor, 1 ml 6-1*H*-UDP-GlcNAc (44,400 dpm/nmol) in 50 mM MES, pH 6.5, in total volumes between 30 and 100 μl. Enzyme sources were nematode microsomal membranes, cell lysates, conditioned media either directly or dialysates against 10 mM MES, pH 6.0, or immunoprecipitates. Assays using microsomal membranes contained 2 μl acceptor and both, which was 10^6^ dpm/ml. In addition these samples contained 5 μM adenosine 5’-monophosphate and 500 μM 2-acetamido-1,2-dideoxyribofuranoside (Toronto Research Chemicals). After 3 h at the appropriate incubation temperature, 1 ml of ice-cold water was added to stop further reaction, and assays were either frozen or processed immediately. Enzyme products were separated from radioactive substrates by binding them to 50 mg of C18 cartridges (Alltech) precondi-

**Fluorescence Analysis of Lec4 Cells Transfected with gly-2—** Transient transfections were performed essentially as above except that
Cel-gly-1 is GlcNAc-TV

Lecl or CHO-K1 cells were plated at 10^4 cells on 6-cm tissue culture dishes (Falcon) that were cotransfected with 0.5 μg of pCMVCD20 (29) and 2.5 μg of pLeu4, pCHO-K1, pCSYK-1, or pCSYK-L116R DNA using 18 μl of LipofectAMINE in 3 ml of OptiMEM-I. At 71 h post-transfection, the media were aspirated, and the plates were rinsed with ice-cold PBS, aspirated, and then filled with 1 ml of PBSE (PBS, 0.1% EDTA, pH 7.4) and dissociated from the dish by incubation in 0.5 ml of PBSE for 10 min at room temperature before triturating with 4.5 ml of PBS, 1% v/v fetal bovine serum, 0.1% w/v Na3PBSE. Aliquots of 1.2 × 10^5 transfected cells were used to 6-ml polypropylene tubes (Falcon) on ice, filled with PBSFN, and centrifuged at 500 g for 5 min at 4°C, and the supernatants were decanted. Fast-conjugated t-PHA (Sigma) was preadsorbed against Lecl4 cells by incubating 40 μg of FITC-t-PHA with 4 × 10^5 untransfected Lecl4 cells (harvested using PBSE) in a total volume of 800 μl of PBSFN for 15 min on ice, then clarified. Lecl4-absorbed FITC-t-PHA (0.5 μg) and 10 μl of phycoerythrin-conjugated monoclonal anti-CD20 (BD PharMingen) were added to each sample, and the cells were resuspended. After a 30-min incubation on ice, tubes were filled with PBSFN and centrifuged at 500 g for 5 min at 4°C, and the supernatants were decanted. Washes were repeated twice more before a final resuspension in 1 ml of PBSFN. FACSc was carried out on a FACStar (BD Pharmingen). Live single cells were selected based on a forward and side scattering gates, and data acquisition and analysis used the CellQuest package. Transfected cells were gated based on the pEGFP-C3, pEGFP-GLY2, pEGFP-L116R transfected cells were assayed immediately for GlcNAc-TV activity.

Biochemicals). After 5 min on ice, lysates were clarified at 14,000 rpm for 5 min in a microcentrifuge, and supernatants were transferred and stored at −70°C. Cell pellets were lysed in 50 mM MES, pH 6.5, 0.5% v/v Triton X-100, 10 mM EDTA, dissociated from the dish by incubation in 0.5 ml of PBSE for 10 min at room temperature before triturating with 4.5 ml of PBS, 1% v/v fetal bovine serum, 0.1% w/v Na3PBSE. Aliquots of 1.2 × 10^5 transfected cells were used to 6-ml polypropylene tubes (Falcon) on ice, filled with PBSFN, and centrifuged at 500 g for 5 min at 4°C, and the supernatants were decanted. Washes were repeated twice more before a final resuspension in 1 ml of PBSFN. FACSc was carried out on a FACStar (BD Pharmingen). Live single cells were selected based on a forward and side scattering gates, and data acquisition and analysis used the CellQuest package. Transfected cells were gated based on the pEGFP-C3, pEGFP-GLY2, pEGFP-L116R transfected cells were assayed immediately for GlcNAc-TV activity.

Transfected cells were further out-crossed with N2 to remove the dpy-14 allele and L. cepacia strain XA728 gly-2 (qa703*)1 and qa703-bearing animals were cross between gly-2 and both marker loci, dpy-5 and unc-13. Of 26 Dpy non-Unc chromosomes, 15 recombinations occurred in the dpy-5-gly-2 interval, and of 28 Unc non-Dpy chromosomes, 13 recombinations occurred between gly-2 and unc-13. Corporation of Precomplementation Lines—The 13,806-bp XbaI fragment corresponding to bases 17,188–30,994 of cosmid C55B7 was subcloned into the XbaI site of pZEO-2 to create pResLNg-9E, and the structure was verified by restriction digests. This genomic region encompassed all gly-2 sequences detected in transscripts as well as an additional 4248 bases upstream of the 5’ limit of yk12688 and 1281 bases downstream of the site of polyadenylation. XA762 hermaphrodites were transformed by gonad injection (30) of a mixture of pResLNg-9E and pRF4, a plasmid containing a region of C. elegans genomic DNA carrying the rol-6(su1006) mutation that acts dominantly to kill animals bearing the array as an X chromosome. Several independent rolling lines were established, and the percentage of rolling self-progeny from each was characterized. GlcNAc-TV activity was assayed using microsomal membranes prepared from 2 such lines, XA766 gly-2(qa703) I; qaEx743[+] (E13G11.1); qa703) I; qaEx745[+]; rol-6(su1006)] and XA768 gly-2(qa703) I; qaEx745[+]; rol-6(su1006)], both of which transmitted the array to 30–50% of their progeny.

Microsomal Membrane Preparation—Cultures were established by picking 50 rolling L4 hermaphrodites (or 20 animals from non-transgenic lines) to each of 5 100-nm diameter complete nematode growth medium plates that were then grown at room temperature until the animals cleared the Escherichia coli OP50 lawn. Nematodes were rinsed from the plates in cold 100 mM NaCl, washed twice, then floated on sucrose (60% w/v). After 2 washes with 100 mM NaCl, the pellet was snap-frozen in an ethanolic-dry ice bath and stored at −70°C. Samples were thawed by adding 1 ml of TSEC (20 mM Tris-HCl, 250 mM sucrose, 10 mM EDTA) and sonicated on ice 5 times using a 10-s pulse before dilution with a further 3 ml of TSEC. After centrifugal clarification for 10 min at 3000 rpm at 4°C (Sorval RT6000), the supernatant was ultracentrifuged at 55,000 rpm for 1 h at 4°C (Beckman L8–80 M with a 70.1Ti rotor). The microsomal pellet was suspended in a minimal volume of 100 mM MES, pH 6.5, 2% v/v Triton X-100, 2 × CompleteTM, 20 mM EDTA, the protein concentration was determined by BCA assay (Pierce) standardized with bovine serum albumin, then 386 μg of each preparation was subjected immediately to GlcNAc-TV assay.

**RESULTS**

The gly-2 Gene of C. elegans—TBLASTN queries of the GenBankTM dbEST data base using rat GlcNAc-TV polypeptide (GB:AAA41665) revealed two homologous C. elegans ESTs,
cm20c4 and yk126h8 (33), which were obtained and sequenced (Fig. 1). A single reverse transcriptase-specific product after 5′ RACE was observed, and direct sequencing revealed a trans-spliced SL1 sequence attached to position 14, where a splice acceptor site occurs immediately upstream in the genomic sequence. Comparison of the genomic and yk126h8 sequences confirms an intron at this point. All 35 independent subclones of the RACE product that were tested for the presence of SL1 and SL2 sequences by colony PCR and 5 that were sequenced contained SL1. This transcript structure is concordant with the Northern analysis that indicated a single poly(A) RNA species of 2.25 kb (Fig. 2). Comparison of the cDNA and genomic sequences indicates that the gene organization is typical, with 10 exons of 82–589 bp separated by 44–882 bp of introns (Fig. 3A) (34). Notably, the majority of the exon boundaries in human and C. elegans genes occurs at equivalent residues, and in most cases, the phase is conserved too. We named the gene gly-2 as a member of the GLYcosylation class. BLAST searches using the cDNA or deduced polypeptide sequences revealed that the C. elegans genome contains a single homologous region, implying that gly-2 is the nematode orthologue of GlcNAc-TV.

The conceptual translation of the open reading frame encodes a 669-amino acid polypeptide that is 59.9% similar and 36.7% identical to rat GlcNAc-TV. When the sequence was queried against GenBank™ using BLAST, only mammalian GlcNAc-TV sequences were returned as significant hits. There are five potential N-linked glycosylation sites, but they are not conserved with the mammalian homologues. Hydropathy plots indicated that GLY-2 is a type II membrane protein with the secondary structural characteristics of Golgi glycosyltransferases (Fig. 4A). This plot reveals four distinct regions in GLY-2; a hydrophilic cytosolic tail precedes the putative TMD, whereas the lumenal part of the molecule consists of a conserved C-terminal portion. Consistent with this model, alignments between GLY-2 and mammalian homologues showed increased conservation in the C-terminal portion of the molecule (Fig. 4B). A conserved peptide (C110–P124) lies in the otherwise diverged stem that encompasses a conserved leucine residue equivalent to that mutated in the GlcNAc-TV gene of Lec4A cells (Fig. 3B) (35).

GLY-2 Has GlcNAc-TV Enzymatic Activity—Based on the

FIG. 1. The sequence of gly-2 cDNAs. A, 5′-untranslated region of EST yk126h8 that is absent in the major gly-2 mRNA. B, cDNA sequence and inferred polypeptide of mature gly-2 mRNA numbered relative to the initiator residues. SL1 is trans-spliced to nucleotide 14. The start of EST cm20c4 is also indicated. Initiator and stop codons are shown in bold. The putative TMD is double-underlined. The consensus polyadenylation signal is boxed. The run of eight adenines preceding the poly(A) (An) indicator is present in the genomic sequence. It is therefore not clear where transcription terminates precisely.

FIG. 2. Northern analysis of poly(A) RNA from C. elegans strains Bristol N2 and CB1490 him-5(e1490) probed with the SalI-SmaI fragment of yk126h8. The data represent one of two independent experiments. The arrows indicate migration positions of size markers and the residual ribosomal RNAs in the preparations. The origins are marked by radioactive pencil lines at the top.
Fig. 3. Genomic structure and alignment detail of gly-2 with mammalian homologues. A, inferred genomic structure of gly-2 by comparison of cDNA and genomic sequence (cosmid C5857, GB:AC006625, top) and human GlcNAc-TV cDNA and genomic sequence (52), GB:D17716 (53) and GB:S80050, bottom. Dashed lines indicate the alignment of the first amino acid of the C. elegans exons with the corresponding aligned residue in the human sequence. The double-headed arrow shows the position of the mislocalization mutations, and the exon containing the conserved stretch around this leucine is shaded light gray. The darker gray exons are those comprising the catalytic domains. The phase of each intron is indicated by the digits 0, 1, or 2. Note that the intron scaling for the human sequence is different. Non-coding exons or portions are unshaded. The regions deleted by alleles qa700 and qa703 are demarcated with “Δ” bars. B, detail of the alignment between Ce-GLY-2 and CHO cell GlcNAc-TV (35) (GB:U62587). Mislocalization is caused by mutation of the emboldened leucine. The island of conservation surrounding this residue is underlined. The triangles indicate the in-frame boundary residues between exons. The arrows indicate the positions of truncations Δ133, Δ137, and Δ138.

Fig. 4. Hydropathy and conservation profiles of the GLY-2 protein sequence. TMD marks the peak corresponding to the TMD, and the solid/dotted line indicates the catalytic portion of the molecule as determined by experiments reported here. The dotted portion of the line reflects that the C-terminal extent of the catalytically critical region was not investigated. A, Kyte-Doolittle profile generated using the GCG Peptidestructure program using a window of 7 residues (51). B, running similarity comparison to rat GlcNAc-TV (GB:L14284) aligned using the GCG Gap program averaging over a window of 25 residues. Alignment position refers to the numbering of the alignment, not a specific residue in either polypeptide. Global similarity is indicated by the dashed line. The asterisk marks the position of the mislocalization point mutations and the surrounding stretch of conserved residues.
Ce-gly-2 Is GlcNAc-TV

expected size (~81 kDa) was observed when the immunoprecipitate of GLY-2ΔΔ27 was Western-blotted for S-tag. In the other truncations an unavoidable background band masked the region at the expected size range (~60 kDa) (Fig. 5F). As with conditioned medium, deleting the first 137 residues of GLY-2, a region comprising the initiator methionine, the TMD, and the predicted stem region, including the C116-P124 peptide, had little effect on specific activity. Removing a single additional residue reduced activity by 75%. Therefore, the boundary of the catalytic domain does indeed correspond to the 5' limit of the exon initiated by Ile-138.

gly-2 Can Rescue the Cell Surface Phenotype of Chinese Hamster Ovary Lec4 Cells—The complementation of a genetic defect by a heterologous allele is a stringent test of equivalence since all the salient properties of the endogenous gene must be fulfilled by the introduced allele in the physiological environment. Lec4 mutant cells lack GlcNAc-TV activity and the mature glycan products, GlcNAcβ1,6 branched N-linked oligosaccharides on cell surface glycoproteins, which can be specifically detected as determinants of 1-PHA binding (Fig. 6A). The parental phenotype was restored to Lec4 by transfecting the wild-type CHO-K1 GlcNAc-TV cDNA expression constructs (Fig. 6B). Transfection with wild-type gly-2 also rescued the Lec4 phenotype, and the profile is qualitatively identical to that of Lec4 cells rescued by transfection of CHO-K1 GlcNAc-TV (Fig. 6D). The partially rescued population is probably the result of low levels of activity expressed in these cells, itself due to thermolability of the nematode enzyme at 30 °C. Thus, gly-2 is functionally equivalent to the mammalian gene product, able to act on the natural glycoprotein substrates found in mammalian cells and create glycans recognized by 1-PHA.

GlcNAc-TV must be present in the medial-Golgi because the elaboration of β6-GlcNAc-branched N-glycans and Lee4A mutant cells cannot bind 1-PHA at the cell surface because they mislocalize active enzyme (35). The equivalent of the Lee4A missense mutation in GLY-2 was assayed. Protein truncations removing this region are catalytically active, yet GLY-2(L116R) failed to rescue the Lec4 phenotype in three independent experiments (Fig. 6C). Thus, although the wild-type GLY-2 enzyme complements Lee4 and, therefore, must be expressed and functional, the L116R mutant might not be. To address this, since attempts to raise anti-GLY-2 antibodies were unsuccessful, as were assays for activity in these transfected samples, constructs expressing GFP fused to the N terminus of GLY-2 were tested. Transfection of pEGFP-C3 alone does not affect the 1-PHA binding properties of Lec4 or CHO-K1 (Fig. 7, A and B). GFP::GLY-2(+) however, results in complete restoration of the parental phenotype in Lec4 cells and is more effective than native GLY-2 (compare Figs. 7D to 6D). Consistent with this enhancement, GFP::GLY-2(L116R) can now partially rescue the cell surface phenotype and must therefore be catalytically competent (compare Figs. 7C to 6C). The FACS analysis indicated that transfection efficiencies were the same for all samples; therefore, cell extracts were Western-blotted for GFP epitopes, and GlcNAc-TV was assayed. Slightly more GFP epitope, as well as GlcNAc-TV enzyme activity, can be detected per cell transfected with GFP::GLY-2(L116R), but there is no indication of appreciable differences in specific activity (Fig. 7,
E and F). Transfected cells were examined by deconvolution microscopy (data not shown), but fluorescent signals from both native and mutant forms were present in membranous compartments other than medial-Golgi. Overexpression by transient transfection may overwhelm retention and trafficking mechanisms, but nevertheless, GLY-2(+ and GLY-2(L116R) have different rescue behaviors.

Expression Pattern of gly-2p::GFP during Nematode Development—Transcriptional fusions of 6.7 kbp of upstream genomic DNA corresponding to bases 19,280–25,991 of cosmids C55B7 to nuclear localized and cytosolic forms of GFP provided by vectors pPD95.69 and pPD95.77, respectively, were used as reporter constructs. This stretch includes the 3‘ end (base 19,436) of the next confirmed gene upstream on the same strand as gly-2. It encompasses all of the 5‘-untranslated region sequences found in yk266h8 (which starts at base 21,436) as well as the region that is conserved in the genome of Caenorhabditis briggsae, a closely related species (alignment starts at base 23,114). By these criteria, the constructs should contain a fully qualified promoter of gly-2.

The distribution of signal in transgenic worms was unique and highly restricted with respect to tissue and/or stage of development but did not correspond to the descendants of a particular branch of the cell lineage. Fluorescence was first detectable at the comma stage (Fig. 8A) in cells that divided and appeared to migrate during the 2-fold (Fig. 8B) and 3-fold stages (Fig. 8C). Neuronal staining was obvious from L1 onward and by early L4 was seen to occur in both the dorsal and ventral nerve chords (Fig. 8D). During this stage, a strong signal was noted in the developing vulva (most likely the vulE and/or vulF cells). By late L4 an intense GFP signal in the spermathecal valve as well as other vulval and/or uterine structures was evident (Fig. 8E). Expression in the uv1 and uv2 cells was suggested by the pattern of fluorescence around the vulva. However, the nuclear-localized reporter construct stained more nuclei than can be accounted for by expression in these cells alone (Fig. 8F). With this construct, nuclear localized signal was observed in all four nuclei of the syncytial spermathecal valve cell (Fig. 8G). Although GFP fluorescence was seen to be strongest in the late L4 and early adult for the spermathecal valve and vulval/uterine structures previously noted, it was seen to persist throughout adulthood (Fig. 8, H–J). The M8 cell of the terminal bulb of the pharynx, all six cells of the pharyngeal-intestinal valve, and neuronal cell bodies within the metacorpus and around the isthmus of the pharynx also expressed gly-2p::GFP (Fig. 8K). At least 37 neurons with cell bodies lying next to the ventral nerve chord were positive for gly-2-directed reporter expression in the adult hermaphrodite, although with widely varying levels of staining. There was also GFP fluorescence present in other neurons associated with the pre-anal, dorso-rectal, and/or lumbar ganglia. In adult males, expression was similar in non-sexually dimorphic tissues and was also observed in axons that project into rays 2, 3, 5, 6, and either 8 or 9 of the copulatory bursa (data not shown).

gly-2 Is a Non-essential Gene—ev581 is a Tc1 insertion
allele into the 7th intron of gly-2 from which qa700 was derived by imprecise excision, an event that deleted 1165 bp containing 2.5 exons that contribute to the catalytic domain (Table I). qa703 is a deletion created by ethylmethanesulfonate-induced deletion mutagenesis that removes 494 bp containing exon 6 and half of the largest exon, 7, both of which contribute to the catalytic domain. Both deletion alleles are probably null, but animals homozygous for either are viable. To check that no gross rearrangements occurred during mutagenesis, genetic mapping of the genotypes was performed. This placed the alleles on linkage group I between 1.07 and 1.18 map units to the right of dpy-5, exactly where expected from interpolations of the physical map.

GlcNAc-TV activity could be detected in microsomal extracts of wild-type C. elegans but not in the deletion mutant strain XA762 gly-2 (qa703) (Fig. 9). Enzyme activity was restored in transgenic lines carrying a genomic region encompassing the gly-2 gene on the deletion mutant background. Thus, gly-2, which is the sole cognate homologue of Mgat-5 in C. elegans, encodes nematode GlcNAc-TV.

The strain XA728 gly-2(qa700**14) I had fertility defects arising from abnormal sperm function (Spe) that were not observed in XA762 gly-2(qa703**10) I. Compound heterozygotes of a qa700/qa703 genotype were non-Spe confirming that this defect is caused by a linked but extragenic mutation in a complementation group unrelated to gly-2 (data not shown). Although gly-2 is expressed in many neurons, the vulva, and spermatheca, XA762 was wild type with respect to morphology, egg laying and hatching, locomotion, brood size, dauer switching,
male incidence, developmental timing, and mechanosensory axon path-finding (data not shown). GFP reporter patterns were also unaffected by the mutant background.

**DISCUSSION**

The genomic structure of the gly-2 gene is significantly related to that of human GlcNAc-TV. The majority of exon boundaries, particularly in the catalytic domain, occur at equivalent residues and are in-frame. The N-terminal boundary of the catalytic region starts at exon 4; exon 3 contains the “Lec4A” region. Retention of phase zero introns in ancient genes is a feature of the “introns-early” model (39). These observations support the notion that exon shuffling of functional domains may have been the mechanism by which the ancestral GlcNAc-TV gene originated.

The deduced polypeptide sequence of gly-2 is stereotypical of Golgi glycosyltransferases, being a type II membrane protein with a 20-residue TMD starting six residues from the N terminus. This length is efficiently retained by the Golgi apparatus and is the sole element in the polypeptide that appears to have bilayer-spanning properties (40). The luminal portion starts with a hydrophilic region that may position the following catalytic domain away from the membrane and so promote efficient interactions with macromolecular substrates. Heterologous expression of recombinant gene product demonstrated that GLY-2 does indeed possess GlcNAc-TV enzyme activity and other properties in common with the mammalian homologue. The putative initiator codon and the TMD were confirmed since soluble recombinant fusion proteins were produced when truncated. The proposed stem could also be removed without affecting GlcNAc-TV enzyme activity in vitro. Several other C. elegans glycosyltransferase-related sequences have been found to possess the catalytic activity expected from their homologies. gly-3, gly-4, and gly-5 are polypeptide GalNAc-Ts (15), and gly-12 and gly-14 encode active GlcNAc-TI (16), whereas CePT-1 is an α1,3-fucosyltransferase (17). gly-1 and possibly the other core 2 GlcNAc-T homologues may be an exception (23). GLY-1 transfers glucose rather than GlcNAc to core 1 acceptors (24), an observation concordant with the available structural data on C. elegans glycoprotein glycans (22). The components of the proteoglycan pathway encoded by sqv-3, sqv-7, and sqv-8 all possess the biochemical activity expected from their homologies (20, 21). The proper functioning of GlcNAc-TV depends not only on catalytic competence but also upon being able to interact with nascent glycoprotein substrates in the endoplasmic reticulum, correct localization, and domain structure (35). We found unequivocally that gly-2 could rescue the surface lectin binding phenotype of Lec4 cells. Thus, GLY-2 retains all of the salient properties of the mammalian GlcNAc-TV despite being diverged for >500Myr (41).

Alignment of mammalian GlcNAc-TV and GLY-2 identified a region that is highly conserved despite being N-terminal to the catalytic domain. This region contains a leucine that is mutated in Lec4A cells, causing otherwise active GlcNAc-TV to mislocalize and fail to elaborate cell surface β-GlcNAc-branched N-glycans in consequence (35). The equivalent mutation in native GLY-2 did not rescue the Lec4 defect, but a GFP fusion product could do so inefficiently. It may be that the fusion protein is better expressed than the native nematode enzyme in Lec4 cells or that the addition of GFP stabilizes the product (42). The simplest interpretation is that GFP::GLY-2(L116R) is mislocalized as in Lec4A, but due to overexpression typical of transient transfections, a portion overwhelms the endoplasmic reticulum retention system and proceeds to the medial-Golgi (43). BLAST searches indicated that the conserved 15-residue peptide encompassing the critical leucine is unique to GlcNAc-TV but has been conserved throughout metazoan radiation. Because mutations affect subcellular localization, it may be that the region is conserved because of a role in targeting to the medial-Golgi. If so, this mechanism is either GlcNAc-TV-specific or acts via its conformational properties, plausible since the peptide is bounded by two conserved cysteines. Conformational elements participate in the subcellular localization of lysosomal hydrolases where a common surface is recognized to initiate formation of the mannose 6-phosphate-targeting signal (44).

Our data are concordant with the dominant transcript being SL1 trans-spliced to the first splice acceptor upstream of the initiator codon and is typical of monocistronic C. elegans genes with a proximal promoter (34). yk126h8 contains an additional 383 nucleotides that occur in 4 non-coding exons 3994–4533 bp upstream and may represent a minor isoform from a distal upstream basal promoter. Distal promoters driving expression of this type of transcript at low levels are observed in C. elegans, for example pbc-1 (45). The genomic fragment used for constructing the GFP reporter transgenes included both potential promoters. From these, GLY-2 expression can be crudely summarized as occurring in some of the structures that have valve properties, the vulva, the spermathecal valve, and the pharyngeal-intestinal valve. The other major locus of expression is neuronal, present in many but not all 302 neurons in the adult hermaphrodite (46). Curiously, mammalian brain is rich in GlcNAc-TV transcripts, but enzyme activity is barely detectable, and Mgat-5° mice are not obviously affected (4). However, failure to nurture pups is significantly more common in Mgat-5° mice in a 129/Sv background.3

The essentially complete sequence of the C. elegans genome (8) contains a single gene that is orthologous to mammalian Golgi GlcNAc-TV proteins at both the primary sequence and domain organization level. This is unusual for glycosylation-related genes in the nematode. The C. elegans genome contains many gene families, and glycosyltransferases are well represented (14, 47, 48). Multiple glycosyltransferase homologues, C-type and S-type lectin domains as well as nucleotide-sugar synthases, occur in a cluster (49). Core 2 GlcNAc-T-like sequences are the 167th largest gene family (23, 48); there are nine polypeptide GalNAc-T-like sequences (15), three homologues of GlcNAc-TI (16), and evidence for at least two α1,3 fucosyltransferases (17). There are two β-galactosyltransferase homologues, of which mutations in one, sqv-3, affects epithelial morphogenesis, resulting in defects in vulval invagination as well as oocyte receptiveness to sperm (18, 19). Many mammalian glycosyltransferases are also present in multiple copies (50), but as in the worm, GlcNAc-TV has only one functional copy. Disruption of the Mgat-5 locus in mice results in a

3. M. Granovsky, J. Pawling, P. Cheung, and J. W. Dennis, unpublished observations.
complete loss of both enzyme activity and GlcNAcβ1,6-branched structures (4). Although structural studies have yet to observe complex N-glycans in C. elegans, GlcNAc-TV activity in wild-type animals is detectable, absent in animals with gly-2 deleted, and restored by transgenes containing gly-2 genomic DNA. From our present study, it appears that Ce-gly-2 is orthologous to Mgat-5, structurally conserved at both genetic and polypeptide levels, and functionally interchangeable with mammalian GlcNAc-TV. Such “deep homology” is a feature of ancient and pivotal genes that occur in conserved pathways (41), but ablation of the gly-2 gene in C. elegans is without visible defects despite resulting in an enzymatically null phenotype. This situation is not unusual; many genes with severely defective alleles are viable in C. elegans (e.g. 23, 54). It may be that the contributions are subtle under laboratory growth conditions. Mgat-5b mice are also without overt phenotype but display several phenotypes that are dependent on extrinsic conditions. Suppression of tumor growth and metastasis induced by the Polyomavirus middle T-antigen is observed (4)m and abnormalities in T-cell function, although significant, do not appear to compromise the animals greatly under laboratory conditions (3). The tractability of screens in C. elegans to uncover synthetic phenotypes enables this conundrum to be addressed and should mutate genes that interact genetically with gly-2. These would reveal GlcNAc-TV-dependent pathways and phenotypes, identifying the contributions to fitness made by β6-GlcNAc-branched N-glycans.

Acknowledgments—We thank Lynda Doughty for DNA sequencing, Giselle Knowles for FACS analyses, and John Hudson and Cindy Todoro for critical readings of the manuscript.

REFERENCES
1. Dennis, J. W., Granovsky, M., and Warren, C. E. (1999) Bioessays 21, 412–421
2. Granovsky, M., Fode, C., Warren, C. E., Campbell, R. M., Martin, J. D., Pierce, M., Fregien, N., and Dennis, J. W. (1995) Glycobiology 5, 797–806
3. Demetriou, M., Granovsky, M., Quaggin, S., and Dennis, J. W. (2001) Nature 409, 733–739
4. Granovsky, M., Fata, J., Pawling, J., Muller, W. J., Khokha, R., and Dennis, J. W. (2000) Nat. Med. 6, 306–312
5. Dennis, J. W., Laferte, S., Waghorne, C., Breitman, M. L., and Kerbel, R. S. (1987) Science 236, 582–585
6. Riddle, D. L., Blumenthal, T., Meyer, B. J., and Press J. R. (eds) (1997) C. elegans II, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
7. Ce-gly-2 Is GlcNAc-TV
8. The Nematode Caenorhabditis elegans
9. Sternberg, P. W., and Han, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 6803–6808
10. de Souza, S. J., Long, M., Klein, R. J., Roy, S., and Gilbert, W. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 6808–6809
11. Fitch, D. H. A., and Thomas, W. K. (1997) in Glycobiology, Vol. 8, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
12. Senior, B. M., Intecho, A., and Porte, S. (1998) Glycobiology 8, 932–934
13. Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biews, M., Bucher, P., Cerutti, C., Corpet, F., Croning, M. D., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., Janssens, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Muller, N. J., Oinn, T. M., Pagini, M., and Servant, F. (2001) Nucleic Acids Res. 29, 37–40
14. Fitch, D. H. A., and Thomas, W. K. (1997) C. elegans II, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
15. Senior, B. M., Intecho, A., and Porte, S. (1998) Glycobiology 8, 932–934
16. Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biews, M., Bucher, P., Cerutti, C., Corpet, F., Croning, M. D., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., Janssens, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Muller, N. J., Oinn, T. M., Pagini, M., and Servant, F. (2001) Nucleic Acids Res. 29, 37–40
17. Mello, C. C., Kramer, J. M., Stinchcomb, D., and Ambros, V. (1991) EMBO J. 10, 3859–3970
18. Liu, X., Spörrle, J. M., Mulligan, E. L., Chen, J., Reardon, B., Westlund, B., Sun, L., Abel, K., Armstrong, B., Huidan, G., King, J., McLaugh, L., Buchan, M., Cleave, R., and Johnson, C. B. (1997) Genome Res. 9, 859–867
19. Waterston, R., Martin, C., Craxton, M., Hyngh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., and Halloran, N. (1992) Nature 358, 114–123
20. Blumenthal, T., and Steward, K. (1997) in C. elegans II, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
21. Senior, B. M., Intecho, A., and Porte, S. (1998) Glycobiology 8, 932–934
22. Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biews, M., Bucher, P., Cerutti, C., Corpet, F., Croning, M. D., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., Janssens, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Muller, N. J., Oinn, T. M., Pagini, M., and Servant, F. (2001) Nucleic Acids Res. 29, 37–40
23. Fitch, D. H. A., and Thomas, W. K. (1997) in C. elegans II, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
24. Senior, B. M., Intecho, A., and Porte, S. (1998) Glycobiology 8, 932–934
25. Fitch, D. H. A., and Thomas, W. K. (1997) in C. elegans II, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
26. Senior, B. M., Intecho, A., and Porte, S. (1998) Glycobiology 8, 932–934
27. Mello, C. C., Kramer, J. M., Stinchcomb, D., and Ambros, V. (1991) EMBO J. 10, 3859–3970
28. Blumenthal, T., and Steward, K. (1997) in C. elegans II, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
29. Senior, B. M., Intecho, A., and Porte, S. (1998) Glycobiology 8, 932–934
The Caenorhabditis elegans Gene, gly-2, Can Rescue the N-Acetylglucosaminyltransferase V Mutation of Lec4 Cells

Charles E. Warren, Aldis Krizus, Peter J. Roy, Joseph G. Culotti and James W. Dennis

J. Biol. Chem. 2002, 277:22829-22838.
doi: 10.1074/jbc.M201390200 originally published online April 5, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M201390200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 47 references, 23 of which can be accessed free at http://www.jbc.org/content/277/25/22829.full.html#ref-list-1