Molecular Cloning and Functional Characterization of a Lepidopteran Insect β4-N-Acetylgalactosaminyltransferase with Broad Substrate Specificity, a Functional Role in Glycoprotein Biosynthesis, and a Potential Functional Role in Glycolipid Biosynthesis*

Nadia Vadaie and Donald L. Jarvis‡

From the Department of Molecular Biology, University of Wyoming, Laramie, Wyoming 82071

Received for publication, May 3, 2004, and in revised form, June 1, 2004
Published, JBC Papers in Press, June 1, 2004, DOI 10.1074/jbc.M404925200

A degenerate PCR approach was used to isolate a lepidopteran insect cDNA encoding a β4-galactosyltransferase family member. The isolation and initial identification of this cDNA was based on bioinformatics, but its identification as a β4-galactosyltransferase family member was experimentally confirmed. The newly identified β4-galactosyltransferase family member had unusually broad donor and acceptor substrate specificities in vitro, as transferred galactose, N-acetylgalactosamine, and N-acetylgalactosamine to carbohydrate, glycoprotein, and glycolipid acceptors. However, the enzyme preferentially utilized N-acetylgalactosamine as the donor for all three acceptors, and its derived amino acid sequence was closely related to a known N-acetylgalactosaminyltransferase. These data suggested that the newly isolated cDNA encodes a β4-N-acetylgalactosaminyltransferase that functions in insect cell glycoprotein biosynthesis, glycolipid biosynthesis, or both. The remainder of this study focused on the role of this enzyme in N-glycoprotein biosynthesis. The results showed that the purified enzyme transferred N-acetylgalactosamine, but no detectable galactose or N-acetylgalactosamine, to a synthetic N-glycan in vitro. The structure of the reaction product was confirmed by chromatographic, mass spectroscopic, and nuclear magnetic resonance analyses. Co-expression of the new cDNA product in insect cells with an N-glycoprotein reporter showed that it transferred N-acetylgalactosamine, but no detectable galactose or N-acetylgalactosamine, to this N-glycoprotein in vivo. Confocal microscopy showed that a GFP-tagged version of the enzyme was localized in the insect cell Golgi apparatus. In summary, this study demonstrated that lepidopteran insect cells encode N-acetylgalactosaminyltransferase that functions in insect cell glycoprotein biosynthesis and that these enzymes function in insect protein N-glycosylation pathways and the means to rigorously examine their functions.

For the past decade, several laboratories have been studying protein N-glycosylation pathways in insect cell systems. This work is important because insects occupy an intermediate evolutionary niche in glyobiology between lower and higher eukaryotes (1–3) and because a better understanding of this pathway is necessary to support the use of insect-based systems for recombinant mammalian glycoprotein production (4–7). Despite being used routinely for recombinant glycoprotein production, however, we still have relatively little basic information on glycoprotein biosynthesis and processing in insect systems to support this important practical application. Thus, our group has focused on the use of a molecular genetic approach to provide an unequivocal view of the processing enzymes that constitute insect N-glycosylation pathways and the means to rigorously examine their functions.

Previously, we molecularly cloned genes encoding class I and class II processing α-mannosidases from a lepidopteran insect cell line and characterized the properties of the gene products in detail (8–13). In parallel, other investigators cloned Drosophila melanogaster genes encoding class I and class II processing α-mannosidases and characterized those gene products, as well (14–17). Together with earlier biochemical studies (18–21), these studies clearly established that insect cells encode and express processing α-mannosidases similar to those of higher eukaryotes and provided new information about how these enzymes function in insect protein N-glycosylation. Similar progress in understanding insect core fucosylation reactions has come from the isolation of a core α-1,3-fucosyltransferase gene from D. melanogaster and functional characterization of the gene product (22), which extended earlier biochemical studies on insect core fucosyltransferase activities (23, 24).

A relatively smaller number of reports have focused on insect genes encoding the glycosyltransferases mediating elongation of N-glycan termini. One reason for this is that insects are generally considered to lack functional levels of these enzymes. Historically, this conclusion has been based on the absence of detectable levels of these enzyme activities in insect cells and the fact that the processed N-glycans of insect-derived glycoproteins are usually trimmed, paucimannose structures consisting of Man₆GlcNAc₂ with or without core fucose residues, rather than elongated, complex N-glycans like those produced by mammalian cells (reviewed in Refs. 1–3 and 25–27). However, more recent reports have shown that lepidopteran insect cell lines actually do have low levels of some of the glycosyltransferase activities mediating elongation of N-glycan termini, including N-acetylgalactosaminyltransferase I (28, 29).

* This work was supported by National Institutes of Health Grant GM49734. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 307-766-4282; Fax: 307-766-5098; E-mail: DLRJarvis@uwyo.edu.

This paper is available on line at http://www.jbc.org
N-acetylglucosaminyltransferase II (28), β-galactosyltransferase (30, 31), and β4-N-acetylgalactosaminyltransferase (30). In addition, various reports have documented the presence of terminal N-acetylgalcosamine, galactose, N-acetylgalactosamine, and even sialic acid residues on N-glycans produced by insect cells (31–43). Finally, putative N-acetylgalcosaminyltransferase, galactosyltransferase, and sialyltransferase genes have been identified in the D. melanogaster data bases (44), and the biochemical functions of the putative fly N-acetylgalcosaminyltransferase I (45) and sialyltransferase (46) gene products have been experimentally confirmed in published studies. These and other relatively recent observations support the hypothesis that insect cells have a branched protein N-glycosylation pathway, which includes a previously unrecognized repertoire of terminal glycosyltransferases that can mediate N-glycan elongation (26, 47).

The goals of the present study were to further test this branched pathway hypothesis and to extend the previous report of van Die et al. (30), who found that insect cell lines derived from Trichopolia ni and, to a lesser extent, Spodoptera frugiperda, have low levels of β4-galactosyltransferase and β4-N-acetylgalactosaminyltransferase activities. This finding was among the first to suggest that insect cells could produce complex N-glycans. It was subsequently extended by demonstrations that a subpopulation of the N-glycans isolated from recombinant glycoproteins produced in T. ni cells have terminal galactose residues (32, 38, 39) and that some N-glycans isolated from uninfected T. ni have terminal N-acetylgalactosamine residues (43). Our specific aims were to molecularly clone a member of the β4-galactosyltransferase gene family from T. ni and examine the function of the gene product. Both goals were successfully accomplished. We isolated a T. ni cDNA that belongs to the β4-galactosyltransferase family and showed that the gene product is closely related to the recently described β4-N-acetylgalactosaminyltransferase of Caenorhabditis elegans (48). The T. ni product has unusually broad donor and acceptor substrate specificities in vitro, but clearly preferred N-acetylgalactosamine as the donor for monosaccharide, glycoprotein, and glycolipid acceptor substrates. These results suggested that the function of the T. ni cDNA product is to transfer N-acetylgalactosamine residues to newly synthesized glycoproteins, glycolipids, or both. This finding had significant implications for the notion that a C. elegans β4-N-acetylgalactosaminyltransferase might be able to participate in both glycoprotein and glycolipid biosynthesis, as will be detailed under “Discussion.” However, in accordance with our interest in insect protein N-glycosylation pathways, we focused the remainder of our present study on the potential role of the T. ni enzyme in N-glycoprotein biosynthesis. The results unequivocally confirmed that the new cDNA isolated in this study encodes a β4-N-acetylgalactosaminyltransferase involved in N-glycoprotein glycan elongation. Thus, this cDNA encodes the enzyme involved in the synthesis of GalNAc-β4GlcNAc-R, also known as LacDiNAc or LDN, which is found in the terminal 4-Man residues (49). Our specific aims were to molecularly clone a member of the β4-N-acetylgalactosaminyltransferase gene to be described in any insect system and only the second experimentally confirmed β4-N-acetylgalactosaminyltransferase gene to be reported in any system.

EXPERIMENTAL PROCEDURES

Cells and Cell Culture—The two insect cell lines used in this study were SF9 (6), which is a subclone of the IPLB-SF21-AE line originally isolated from S. frugiperda ovaries (51), and Tn-5B1-4, also known as High Five™ (Invitrogen), which was originally isolated from T. ni embryos (52). Both cell lines were routinely maintained as suspension cultures at densities of about 0.3–3.0 × 10⁶ cells per ml in TXM-PH medium (6) supplemented with 10% (v/v) fetal bovine serum (Hyclone Inc., Logan, UT) and 0.1% (v/v) pluronic F68 (BASF Wyandotte Corp., Parsippany, NJ; Ref. 53). This same medium was also used to culture Tn-5B1-4 cells as adherent cultures in 25 cm² flasks (Corning Inc., Corning, NY) containing about 1.5–6.0 × 10⁶ total cells. In addition, a separate culture of SF9 cells was maintained in suspension at densities of about 0.3–3.0 × 10⁶ cells per ml SFX-INSECT serum-free medium (Hyclone) for use in the enzyme purification experiments.

Preparation of cDNA from Membrane-bound T. ni RNA—Microsomal membranes were isolated from log phase Tn-5B1-4 cell cultures using a previously described method (54). Briefly, about 1 × 10⁶ cells were washed with ice-cold Tris-buffered saline (50 mm Tris-HCl, pH 7.0, containing 0.9% w/v NaCl), then the cells were resuspended in 1 mM MgCl₂ and Dounce-homogenized on ice until most of the cells appeared to be broken under a phase contrast microscope. The homogenates were centrifuged for 5 min at 10,000 × g at 4 °C, then the supernatants were harvested and mixed with an equal volume of 60% (w/v) sucrose in membrane buffer (1 mM Tris-HCl, pH 7.5 containing 1 mM MgCl₂). 10-ml aliquots of this solution were then applied to hand-layered step gradients in membrane buffer, which consisted of 3% of 60% (w/v) sucrose, 7 ml of 45% (w/v) sucrose, and 7 ml of 40% (w/v) sucrose. The samples were overlaid with 7 ml of 25% sucrose, and then the gradients were centrifuged for 20 h at 20,500 rpm at 4 °C in a Beckman SW28 rotor. The membrane band at the 45–60% sucrose interface was collected by side puncture, diluted 1:5 with membrane buffer, and pelleted for 1 h at 29,000 rpm at 4 °C in a Beckman Ti45 rotor. The pellets were resuspended in TE buffer (0.1 mM Tris-HCl, pH 8.0, containing 1 mM EDTA), and the membrane-bound RNA was extracted with phenol, phenol-chloroform (1:1), and then ethanol-precipitated, redissolved, and quantified by spectrophotometry. The RNA was further purified by oligo dT-cellulose column chromatography, as previously described (55), then converted to cDNA with the GeneRacer™ kit (Invitrogen), according to the manufacturer’s protocol, with GeneRacer™ RNA Oligo as the 5’ anchor and GeneRacer™ Oligo ωT as the reverse transcription primer.

Cloning and Functional Analysis of Tn4GalNAcT

Isolation of a T. ni cDNA Encoding a β4-Galactosyltransferase Family Member—The cDNA produced with the membrane-bound T. ni RNA was subsequently used as the template for nested PCR (Ref. 56) with degenerate oligonucleotide primers designed against conserved regions of known β4-galactosyltransferase family members. The sequences of the conserved regions and degenerate primers that ultimately yielded significant PCR products are shown in Table I. The sequences were performed in a total volume of 0.05 ml and contained 10 mM Tris-HCl (pH 9.0), 0.1% (v/v) Triton-X-100, 50 mM KCl, 1.5 mM MgCl₂, 200 μM dNTPs, 1.25 unit of Tag polymerase (Promega), 40 μM each primer, and 100 ng of cDNA. These reactions were incubated for 2 min at 95 °C prior to addition of the primers, then the primers were added and the reactions were cycled 15 times using: (i) 2 min at 95 °C, (ii) 25 s at 55 °C, and (iii) 1 min at 72 °C in a PerkinElmer Applied Biosystems Model 2400 GeneAmp thermal cycler (Foster City, CA). A touchdown PCR method (57) was used in which the annealing temperature, which was 55 °C in the first cycle, was reduced by 1 °C per cycle during these first 15 cycles, and then 35 additional cycles were performed using: (i) 30 s at 95 °C, (ii) 20 s at 40 °C, and (iii) 1 min at 72 °C. After a final extension period of 6 min at 72 °C, aliquots of the spent reactions were used as the templates for secondary PCRs under the same conditions as for the primary PCRs. The spent secondary PCRs were harvested and analyzed on 1% Seaplaque™ agarose gels. A specific amplification product of about 135 bp was resuspended in the gel, cloned into pCR2.1-TOPO® (Invitrogen) and sequenced using the primers. The resulting nucleotide sequence was used to design an internal, gene-specific primer, which was used for nested 3’-RACE1 with the

Footnotes:
1 The abbreviations used are: RACE, rapid amplification of cDNA ends; GFP, green fluorescent protein; Gn Gn-PA, [GlcNAc1-Man6j(GlcNAc1-Man3jMan0jGlcNAc1jGlcNAc labeled with pyridylamine at its reducing terminus; GST, glutathione S-transferase;
GeneRacer™ kit. The primary reactions utilized a gene-specific sense primer (5’-TCGACACGAGGAACTTCTC-3’) and the antisense GeneRacer™ 3’ primer, while the secondary reactions utilized a downstream, gene-specific sense primer (5’-CCCGAGATATGCTTCTCGC-3’) and the antisense GeneRacer™ nested 3’ primer. The template for the primary reactions was 100 ng of the T. ni cDNA preparation described above and the template for the secondary reactions was an aliquot of the spent primary reaction. In both cases, the 3’-RACE reactions were incubated for 2 min at 94 °C prior to addition of the primers, then the primers were added and the reactions were cycled according to the manufacturer’s protocol, which included five cycles of (i) 30 s at 94 °C, (ii) 1 min at 72 °C, and then five cycles of (i) 30 s at 94 °C, (ii) 30 s at 70 °C, (iii) 1 min at 72 °C, and then 20 cycles of (i) 30 s at 94 °C, (ii) 30 s at 50 °C, (iii) 1 min at 72 °C. After a final extension period of 10 min at 72 °C, the spent secondary 3’-RACE reactions were harvested, analyzed by agarose gel electrophoresis, as described above, and a specific amplification product of about 210 bp was recovered from the gel, cloned into pCR2.1-TOPO (Invitrogen), and sequenced using universal primers. The resulting sequence data were used to design two pairs of additional gene-specific primers, which were screened in a λUNZAP (Stratagene) T. ni cDNA library by sibling selection and PCR, as described previously (8, 9). This library was derived from poly(A)+ RNA isolated from Tr-368 cells (58) and was kindly provided by Dr. Paul Friesen of the University of Wisconsin. Ultimately, a specific lambda clone was identified, plaque-purified, and the plasmid insert was excised using the ExAssist® method (Stratagene) according to the manufacturer’s protocol. Bacterial colonies containing this plasmid were identified, amplified, and the plasmid, designated pTnβGalAT, was extracted, purified, and used to sequence both strands of the entire cDNA insert with universal and gene-specific primers. The cDNA sequence was assembled from the raw nucleotide sequence data using GenBank™ as accession number AY601103. This sequence was translated and multiple sequence alignments performed using SIXFRAME and ClustalW, respectively, via the San Diego Supercomputing Center Biology Workbench website (workbench.sdsc.edu/). Biology Workbench and ClustalW, respectively, via the San Diego Supercomputing Center were utilized and multiple sequence alignments performed using SIXFRAME to identify related sequences in the nonredundant protein database.

Expression of the T. ni cDNA with an Immediate Early Baculovirus Vector—The putative T. ni β-galactosyltransferase cDNA was excised by digesting pTnβGalAT with BamHI and DraI and then the 2179-bp product was gel-purified and subcloned into the downstream of the putative baculovirus gene-specific primer of the putative promoter (GAFF-RTACGCT-3’), which added a sequence encoding the immediate early baculovirus transfer plasmid, pAcPv−/JEVT (63). This resulting plasmid, designated pAcPv−/JEVTβGalAT, encoded the full-length putative β-galactosyltransferase family member under the transcriptional control of a baculovirus ie1 promoter and hr5 enhancer. This plasmid was purified by a standard CsCl/ethidium bromide density gradient centrifugation method (62), mixed with BaeIII-digested BakPak6 baculoviral DNA (63), and the mixture was used to co-transfect Sf9 cells by a modified calcium phosphate precipitation method (6). The baculovirus progeny were resolved by plaque purification and an immediate early recombinant was identified by its occlusion positive/white plaque-globoseries phenotype and subjected to two additional rounds of plaque purification. Working stocks of this virus, designated AcPv−/JEVTβGalAT, were prepared by low multiplicity infection of Sf9 cells, titered by plaque assays in the same cell line, and used to express the T. ni cDNA product. Briefly, Sf9 cells were infected with either AcPv−/JEVTβGalAT or wild-type baculovirus at a multiplicity of infection of 5 plaque-forming units per cell. At 24 h post-infection, the infected cells were harvested, washed once with 75 mM sodium cacodylate, pH 7.0, and extracted for 10 min on ice with β-galactosyltransferase buffer (50 mM sodium cacodylate, pH 7.0; 20 mM MnCl2; 1.0% (v/v) Triton X-100; 4 mM ATP, and Complete™ protease inhibitor mixture (EDTA-free, Roche Applied Science, Indianapolis, IN)). The cell extracts were clarified for 10 min at top speed in a refrigerated micro-centrifuge (Hermle Model Z180M), the supernatants were harvested, and total protein concentrations were determined using a commercial bicinchoninic assay kit (Pierce).

β-Galactosyltransferase Family Member Assays—Triplicate samples of infected cell extracts containing 50 μg of total protein were assayed in a total reaction volume of 0.05 ml containing 1 mM p-nitrophenyl N-acetyl-β-D-glucosaminide (pNP-GlcNAc) as the acceptor substrate and 0.3 μCi of UDP-galactose, [1-3H(N)]glactose (American Radiolabeled Chemicals, Inc. St. Louis, MO; 9.1 Ci/mmol), 0.3 μCi of UDP-N-acetyl-D-glucosamine (American Radiolabeled Chemicals, 36 Ci/mmol), or 0.2 μCi of UDP-N-acetylgalactosamine, [8-3H(N)]galactose (American Radiolabeled Chemicals, 15 Ci/mmol) as the donor substrate. The reactions were incubated for 1 h at 37 °C and then terminated by adding 1.0 ml of ice-cold water. The spent reactions were then immediately applied to Sep-Pak C18 cartridges (Millipore Corp., Bedford, MA), which were subsequently washed with water and eluted with methanol. The eluants were added to 5 ml of Scisintase Plus 50% scintillation mixture (Fisher Scientific), and incorporated radioactivity was measured in a Model LS-6500 liquid scintillation spectrometer (Beckman Coulter Instruments). These same assay conditions were used to measure the activity of a purified affinity-purified β-galactosyltransferase protein (shown diagramatically in Fig. 5A), which is described below. We also measured the ability of the purified enzyme to transfer each of the three donor substrates to ovalbumin with a trichloroacetic acid precipitation assay, as described previously (64). Finally, we assayed the ability of the purified enzyme to transfer each of the three donor substrates to the glycopilid acceptor, β-D-galactosyl-acetylgalactosamine, using the Sep-Pak method described above. The results of all of these assays were presented as the average amounts of monosaccharide transferred/μg protein with standard deviations calculated from the individual measurements.

Expression and Purification of an Affinity-tagged, Soluble Form of the T. ni cDNA Product—The nucleotide sequence encoding the predicted soluble domain of the putative T. ni β-galactosyltransferase family member (amino acids 33-421; Fig. 1) was amplified using standard PCR conditions (56) with pTnβGalAT as the template and sense (5’-GAAATCATGAGAAACATTGGGCGCCTTTGATTGACGGC-3’) and antisense (5’-GAATTCCCTATCGACTACCGCCTCGTCTC-3’) primers, which added a sequence encoding the 12 amino acid HPC4 epitope (66) onto the 5’-end and EcoRI sites onto both ends of the product. The PCR product was gel-purified, cloned into pCR2.1-TOPO®, and a clone with the correct nucleotide sequence was identified by sequencing with universal primers. The EcoRI fragment was then excised, gel-purified, and subcloned into the unique site of pTR-BakPak6 (61), which transferred the upstream GST coding sequence had been deleted from EcoRI to BamH1. Sequencing revealed that the insert was out-of-frame with respect to the upstream sequence encoding the baculovirus gp64 signal peptide. This problem was solved by partially digesting the plasmid with EcoRI and religating it, as the correct reading frame was verified by sequencing the resulting plasmid. This final plasmid was used to isolate a recombinant baculovirus designated AcHPC4sTnβGalAT in Sf9 cells, as described above, except this virus was identified by its occlusion-negative, white plaque phenotype, and it encodes the HPC4-tagged predicted soluble domain of the T. ni cDNA product (shown diagramatically in Fig. 5A) under the control of the strong polyhedrin promoter. The plasmid was amplified as described, and working stocks were used to infect 50 ml of Sf9 cells cultured in suspension to a density of about 1 × 10^7 cells/ml in Sf9-INSECT serum-free medium (HyClone) at a multiplicity of infection of 5 plaque-forming units per cell. The infected cells were harvested at 72 h post-infection, and cell-free medium was prepared by low-speed centrifugation. The cell-free medium was adjusted to 3 mM EDTA to inhibit trypsin activity, and proteinase inhibitor cocktail (AP3Cer; Ref. 65), using the Sep-Pak method described above. The results of all of these assays were presented as the average amounts of monosaccharide transferred/μg protein with standard deviations calculated from the individual measurements.

Characterization of the Purified, Soluble Form of the T. ni cDNA Product—The affinity-purified fusion protein described above was initially analyzed by discontinuous SDS-PAGE (67) with Coomassie Blue staining. In parallel, samples were analyzed by SDS-PAGE and immu-
Cloning and Functional Analysis of Tnβ4GalNAcT

The glycoprotein used for these studies was GST-SiManI, a GST-tagged soluble domain of the S. frugiperda class I Golgi α-mannosidase (8), which was expressed under the control of the strong polyhedrin promoter by infecting Sf9 cells with a recombinant baculovirus AcGST-SiManI (10). Three 50-ml Sf9 suspension cultures were grown to a density of about 1 x 10⁹ cells/ml in TNM-FH and then simultaneously infected at a total multiplicity of infection of 10 plaque-forming units per cell for each virus with mixtures of AcGSTSf-ManI plus wild-type baculovirus, AcPc+Ie4/GalAT (71), or AcPc+Ie4/αgal4/GalAT. The infected cells were harvested at 72 h post-infection and the GST-SiManI was affinity-purified, as described previously. Samples of the purified fusion protein were then analyzed by SDS-PAGE with immunoblotting or lectin blotting, as described previously (73). Parallel samples of the model glycoprotein isolated from each infected cell type were treated with buffer alone or PNGase-F (New England Biolabs, Beverly, MA) to demonstrate the specificity of the lectin blotting assay, as described previously (72, 73). The primary and secondary antisera used for the immunoblotting analyses were rabbit polyclonal anti-GST (Sigma) and alkaline phosphatase-conjugated goat anti-rabbit IgG (Sigma), respectively. The lectins used in this study were biotinylated Ricinus communis agglutinin (RCA; EY Laboratories, San Mateo, CA), Wisteria floribunda agglutinin (WFA; Vector Laboratories, Burlingame, CA), and Griffonia simplicifolia lectin II (GS-II; Vector), with alkaline phosphatase-conjugated streptavidin (Vector) as the secondary detection reagent. Both types of assay were performed using a standard alkaline phosphatase color reaction (69).

Intracellular Distribution of the T. ni cDNA Product—A DNA fragment encoding the full-length putative T. ni β4-galactosyltransferase family member (amino acids 1–421) was produced by PCR under standard conditions (56) with pTnβ4GalAT as the template and sense 5’-CCATAGGCAGCCGCGTC-3’ and antisense 5’-GAATCCCGCTACGCTCGTCAATGTTCACTAGTAT-3’ primers that added unique BglII and BamHI sites to its 5’- and 3’-ends, respectively. This fragment was gel-purified, cloned into pCR2.1-TOPO, and a clone with the correct nucleotide sequence was identified by sequencing with universal primers. The 1290 bp BglII-BamHI fragment was cloned into pTnβ4GalAT and sequenced, and then the affinity-purified fusion protein (GFP; Ref. 74). The fused coding sequence was subsequently excised with BglII and NotI, gel-purified, and subcloned into the corresponding sites of the immediate early baculovirus transfer plasmid, pEGFP-N1 (BD Biosciences Clontech), which positioned the full-length T. ni protein coding sequence upstream and in-frame with a sequence encoding the red-shifted variant of Aequorium victoria green fluorescent protein (GFP; Ref. 74). The fused coding sequence was subsequently excised with BglII and NotI, gel-purified, and subcloned into the corresponding sites of the immediate early baculovirus transfer plasmid, pAcPc+Ie4/αgal4/GalAT-GFP, as described above. This virus was identified by its occlusion-positive, white plaque phenotype and purified, amplified, and titered as described above. The intracellular distribution of the fusion protein was determined in Sf9 cells infected with the recombinant baculovirus and harvested at 72 h post-infection. The infected cells were harvested and washed with fresh medium at 24 h post-infection. The infected cells were stained with 1 μg/ml anti-GST (BD Biosciences Clontech) and goat anti-mouse IgG (Chemicon, CA) followed by 1 μg/ml FITC-conjugated goat anti-rabbit IgG (Jackson ImmunoResearch, PA) and then examined under a Leica TCS SP2 confocal laser-scanning microscope (Leica Microsystems, Heidelberg, Germany).

RESULTS

Isolation and Characterization of an Insect β4-Galactosyltransferase Family Member—The CDNA described in this study was isolated on the basis of short, conserved amino acid sequences in a β4-galactosyltransferase family members, which were noted in the literature by several different investigators during the late 1990s (75–80). Degenerate oligonucleotides encoding four of these conserved amino acid sequences (Table I) were synthesized and used as primers for nested PCRs with CDNA prepared from T. ni membrane-bound RNA, as described under “Experimental Procedures.” No specific amplification products were observed upon analysis of the primary PCRs by agarose gel electrophoresis and ethidium bromide staining (primers S6 and A5; data not shown). However, a specific product of about the expected size (135 bp) was observed in the secondary, nested PCRs performed with primers S4 and A6 (data not shown). This PCR product was cloned, sequenced,
Conserved amino acid sequences are shown in lower case, three-letter code, with the central line indicating the most highly conserved sequences and some common substitutions on top. The degenerate oligonucleotides used in this study are shown in upper case below their corresponding sequence. The degenerate nucleotide sequences are shown.

| Name | Sequences |
|------|-----------|
| val | ala ile ile ile pro phe arg |
| AA1 | val |
| S6  | 5'- GCI ATI ATX ATX CCX TTT C-3' |
| AA2 | trp gly gly glu asp asp asp |
| Target | 5'- TGG GGX GGX GAG GAT GAT GAT-3' |
| A5  | 3'- ACC CCX CCX CTC CTA CTA CTA-5' |
| AA3 | his asp val asp leu leu pro |
| S4  | 5' CAT GAT GTI GAT CTX CTX CC-3' |
| AA4 | phe gly gly val ser ala |
| Target | 5'- TTT GGX GGX GTX TCX AGI GC-3' |
| A6  | 3'- AAA CCX CCX CAI TC C-5' |

analyzed by BLAST-P (60), and the results indicated that it was most closely related to β4-galactosyltransferase family members (data not shown). Thus, the sequence of this PCR product was used to design a gene-specific primer for 3'-RACE, as described under “Experimental Procedures.” The 3'-RACE reactions yielded a specific amplification product of about 210 bp, which was also cloned and sequenced. The resulting sequence was used to design exact-match primers, which were used to screen a TN-368 cell cDNA library by a sibling selection method, as described previously (8, 9). This screen ultimately yielded a positive lambda clone, and the cDNA insert was excised and sequenced, as described under “Experimental Procedures.”

The nucleotide and deduced amino acid sequences of this new T. ni cDNA are shown in Fig. 1 and have been deposited in GenBank™ (acc. no. AY601103). The 1266 nucleotide open reading frame encodes a theoretical protein of 421 amino acids with a calculated molecular weight of 48,305. Computer analysis of the putative amino acid sequence with various algorithms indicated that it includes a single transmembrane domain near the N terminus. For example, the TMHMM algorithm predicted a single transmembrane domain from amino acids 7 to 29 of the putative protein. Further analysis using NetNGlyc (59) revealed that the putative protein has five consensus N-glycosylation sites. BLAST-P analysis revealed that the putative cDNA product is most similar to known members of the β4-galactosyltransferase family (Table II). The most statistically significant putative β4-galactosyltransferase family members from Anopheles gambiae and D. melanogaster, as well as a known β4-N-acetylglucosaminyltransferase from C. elegans (48). The specific similarities between the putative T. ni β4-galactosyltransferase family member and selected known or predicted β4-galactosyltransferase family members are shown in a ClustalW multiple sequence alignment (Fig. 2). The key features conserved among the members of this family include a metal binding site (DXD), four cysteine residues, and the major sequence motif WGGGEDDD (76–78, 80, 81). The deduced T. ni cDNA product includes the metal binding site (amino acids 274–276), three of the four conserved cysteine residues (amino acids 196, 289, and 288), and the WGGGEDDD motif (amino acids 334–342; Fig. 2). However, the sequence of the metal binding site is DGD, rather than DVD, which is found in most other family members. A phylogenetic tree (82) clearly revealed that the putative T. ni β4-galactosyltransferase family member is most closely related to the predicted insect and known worm enzymes (Fig. 3). Together, the results of these in silico analyses strongly suggested that the new T. ni cDNA isolated in this study encodes a member of the β4-galactosyltransferase family. In addition, the close relationship of the derived cDNA product to the recently characterized C. elegans family member (48) indicated that this product might be a β4-N-acetylglucosaminyltransferase.

Expression and Biochemical Characterization of the Native T. ni cDNA Product—The enzymatic activity of the new T. ni cDNA product was initially examined in crude lysates of SF9 cells infected with a baculovirus expression vector designed to express the full-length, native gene product during the immediate early phase of infection. This expression strategy was chosen because the specific activity of intracellular forms of some recombinant secretory pathway enzymes are higher when expressed earlier in infection with this type of baculovirus vector, as compared with a conventional baculovirus vector (61). Lysates from SF9 cells infected in parallel with a wild-type baculovirus were used to measure the background in each assay, which was subtracted from the levels measured with the recombinant virus-infected cell lysates. The results of assays with UDP-galactose, UDP-N-acetylgalactosamine, or UDP-N-acetylglucosamine as the donor substrates and pNP-GlcNAc as the acceptor substrate are shown in Fig. 4. In each case, the
glycosyltransferase activities measured in the recombinant virus-infected lysates were above background and there was a clear preference for UDP-N-acetylgalactosamine over UDP-galactose and UDP-N-acetylglucosamine. These results provided preliminary evidence that the \textit{T. ni} cDNA isolated in this study encodes a functional \(\beta\)-galactosyltransferase family member that can preferentially transfer N-acetylgalactosamine to a carbohydrate acceptor.

**Expression, Purification, and Biochemical Characterization of a Tagged, Soluble Form of the \textit{T. ni} cDNA Product**

A solubility figure showing the nucleotide and derived amino acid sequences of the putative \(\text{T. ni}\) \(\beta\)-galactosyltransferase family member. The 5' and 3' untranslated regions are indicated by lowercase letters, a predicted transmembrane domain near the N terminus of the derived protein is underlined, and the consensus N-glycosylation sites are double-underlined.

**FIG. 1.** Nucleotide and amino acid sequences of a putative insect \(\beta\)-galactosyltransferase family member.
ble, HPC4-tagged form of the cDNA product, which is shown diagrammatically in Fig. 5A, was used to confirm and further characterize its enzyme activity. A fragment encoding the predicted soluble domain of the cDNA product was PCR-amplified, fused to an HPC4-epitope coding sequence, and the chimeric sequence was transferred into a conventional recombinant baculovirus. This virus was used to express the fusion protein under the control of the strong polyhedrin promoter in Sf9 cells and then the secreted product was immunoaffinity-purified, as described under “Experimental Procedures.” This conventional baculovirus-based expression strategy was used for these experiments because the polyhedrin promoter provides higher levels of foreign gene expression than the other baculovirus promoters.

The results showed that the purified enzyme preferentially transferred N-acetylgalactosamine to the synthetic N-glycan acceptor and suggested that the enzyme produced a terminally di-N-acetylgalactosaminylated N-glycan through one or both terminally mono-N-acetylgalactosaminylated intermediates. This interpretation was confirmed by MALDI-TOF analysis of the material in the first and second peaks observed in the HPLC profile in Fig. 8D. Because it was necessary to perform scaled-up, incomplete glycosyltransferase reactions to obtain adequate amounts of both the intermediate and end product for the mass spectroscopic analyses, the GnGn-PA acceptor, putative intermediate, and putative end-product were not resolved in the preparative HPLC runs (data not shown). Nevertheless, MALDI-TOF showed that one of the two major species in the first peak from the HPLC run had a mass of about 1620 daltons (Fig. 9A), which was nearly identical to the theoretical mass of a sodium adduct of the presumed intermediate, consisting of GnGn-PA with one terminal N-acetylgalactosamine residue (1621.3 daltons). The other major peak in this profile had a mass of about 1417 daltons, which was nearly identical to the mass of a sodium adduct of the acceptor substrate (1418.3 daltons). Similarly, MALDI-TOF showed that one of the major species in the second peak from the HPLC run had a mass of about 1823 daltons (Fig. 9B), which was nearly identical to the theoretical mass of a sodium adduct of the presumed end-product, GnGn-PA with two terminal N-acetylgalactosamine residues (1824.3 daltons). The other major peak in this profile had a mass nearly identical to the expected mass of the intermediate. The additional minor peaks preceding the major peaks differ by the size of a single sodium ion and those following the major peaks are presumed to be isotopic hydrogen variants, as they differ by a single dalton.

The nature of the terminal N-acetylgalactosamine linkage to GnGn-PA was initially examined by digesting the products of the transferase reaction with β-N-acetylgalactosaminidase prior to HPLC analysis. The results showed that β-N-acetylgalactosaminidase treatment (Fig. 10C) eliminated both of the reaction products observed in the buffer control (Fig. 10B) and produced a single peak that comigrated with the trimannosyl core standard, Manα1→3GlcNAc2-PA (Fig. 10C). Thus, treatment with this exoglycosidase removed the terminal N-acetylgalactosamine and penultimate N-acetylgalactosamine residues from the reaction products, indicating that the T. ni enzyme trans-
ferred terminal N-acetylgalactosamine residues to GnGN-PA in β-linkages. This conclusion was directly confirmed and extended by 500-MHz $^1$H NMR analysis. The NMR spectrum indicated that the sample contains a bi-antennary N-type oligosaccharide ending in PA-derivatized N,N-diacyetylchitobiose (Fig. 11). The bi-antennary nature of this glycan was evident.

![Clustal-W comparison of a putative T. ni and known β4-galactosyltransferase family members.](http://www.jbc.org/)

| Family       | Accession  | Sequence (partial)                                                                 |
|--------------|------------|-----------------------------------------------------------------------------------|
| T. ni        | Tn4GalNAcT | MLFLKANLDPGRAGAICLXYYLVFVGPRSDGQSSA - TSLKLSTRTTAVHATBYGNASDQAGSS                 |
| Hu4GalNAcT   |            | MFWRLAVKQDGKLCAVELLVH - AMEYKPSRTMNNIGGSTLT                                      |
| Hu4GalIT     |            | MRLGKPPCGLIALMGDDGYASQDGRGPOSSASS - DLRSLQCGTVSPLGQSLNGS                        |
| Hu4GalITV    |            | MLRLGKPPCGLIALMGDDGYASQDGRGPOSSASS - DLRSLQCGTVSPLGQSLNGS                        |
| Hu4GalITV    |            | MLRLGKPPCGLIALMGDDGYASQDGRGPOSSASS - DLRSLQCGTVSPLGQSLNGS                        |

**Fig. 2. Clustal-W comparison of a putative T. ni and known β4-galactosyltransferase family members.** This figure shows a Clustal-W (104) multiple sequence alignment of the derived T. ni protein (Tn4GalNAcT), six known human β4-galactosyltransferases (Huβ4GalT-I to VI), predicted D. melanogaster (Dm CG8536), and A. gambiae β4-galactosyltransferases, and a known C. elegans β4-N-acetylgalactosaminyltransferase (Ceβ4GalNAcT). The uppercase white letters in black boxes indicate amino acid identities among all the proteins, the uppercase white letters in gray boxes indicate amino acid identities among some of the proteins, and the unboxed black letters with no boxes indicate amino acid differences.
from the set of chemical shifts of the Man H-1 and H-2 atoms (83). Furthermore, both branches of the bi-antennary oligosaccharide terminated with N-acetylgalactosamine residues \( \beta(1\rightarrow4) \)-linked to N-acetylgalactosamine, as seen by the pair of \( \beta\)-galactosyltransferase family members, the postulated in vivo 4-galactosyltransferase family member, Tn GalNacT). The opposite results were obtained in lectin blots of a model glycoprotein was affinity-purified from each infected Sf9 cell. The protein was analyzed by SDS-PAGE with Coomassie Blue staining (CB) and by immunoblotting with anti-protein C (Ab). The positions of molecular mass standards are indicated by their molecular masses in kDa, on the left-hand side of the figure.

**In Vivo Function of Tnβ4GalNAcT**—The postulated in vivo role of Tnβ4GalNAcT in N-glycoprotein biosynthesis was examined by co-expressing the full-length, untagged enzyme together with a secreted N-glycoprotein, GST-SfManI, which we have used as a reporter in several previous insect N-glycan processing studies (10, 72, 85, 86). Sf9 cells were co-infected with a conventional baculovirus encoding GST-SfManI plus an immediate early recombinant baculovirus encoding either Tnβ4GalNAcT- or bovine β4GalT-I as a control, and then the model glycoprotein was affinity-purified from each infected culture and analyzed by lectin blotting, as described under “Experimental Procedures.” These experiments were specifically designed to include a temporal difference in the expression of the modifying enzymes and the reporter glycoprotein. The \( \beta\)-4-galactosyltransferase family members had to be expressed early in infection so each would be available intracellularly prior to expression of the reporter, which occurred later, under polyhedrin control. The results of these analyses clearly showed that the GST-SfManI produced by cells co-infected with AcP(+)IEtNβ4GalNAcT reacted with WFA, a lectin specific for \( \beta\)-4-linked N-acetylgalactosamine residues (Fig. 12B), but not with RCA, a lectin specific for \( \beta\)-4-linked galactose residues (Fig. 12C). The opposite results were obtained in lectin blots of the GST-SfManI produced by cells co-infected with the control virus, AcP(+)IEβ4GalT. Standard immunoblotting analysis with polyclonal anti-GST showed that approximately equal amounts of GST-SfManI were loaded in each lane (Fig. 12A). The specificity of the lectin blotting assays was verified by pre-digesting the GST-SfManI with PNGase-F (Fig. 12, lanes marked +), which precluded any detectable lectin binding. We also found that GST-SfManII produced by cells co-infected with AcP(+)IEtNβ4GalNAcT failed to bind to GS-II, a lectin specific for terminal N-acetylgalactosamine residues (data not shown). In
contrast, this lectin bound to a control GST-SfManII preparation isolated from cells coinfected with AcP(+/H11001)IEHGnTI, which expresses human N-acetylglucosaminyltransferase I (87). Thus, these results demonstrated that Tn/H92524GalNAcT can participate in N-glycoprotein biosynthesis in vivo and that it transfers N-acetylgalactosamine, but no detectable galactose or N-acetylglucosamine, to the reporter N-glycoprotein used in this study.

Intracellular Distribution of Tn/H92524GalNAcT—If Tn/H92524GalNAcT participates in the elongation of N-glycoprotein glycans, it would be expected to perform this function in the Golgi apparatus. Thus, the intracellular distribution of Tn/H92524GalNAcT was examined in living cell co-localization experiments. The full-length Tn/H92524GalNAcT coding sequence was positioned in-frame with a downstream GFP coding sequence and the resulting construct was used to produce an immediate early recombinant baculovirus. This type of baculovirus vector provides relatively low expression levels, which avoids potential localization artifacts, and allows the recombinant protein to be visualized earlier in infection, prior to the onset of adverse effects of baculovirus infection on host secretory pathway function (88). Sf9 cells were infected for 24 h with AcP(+/HEHnTGnTI)IEHGnTI-FP, counterstained with BODIPY®TR ceramide, and examined by confocal laser scanning microscopy, as described under “Experimental Procedures.” The results showed that the fusion protein had a punctate distribution throughout the cytoplasm in these unfixed, living cells (Fig. 13A). This pattern overlapped almost perfectly with the staining pattern obtained with BODIPY®TR ceramide (Fig. 13C, merged with GFP pattern in Fig. 13B). BODIPY®TR ceramide is a Golgi-specific dye that preferentially stains the trans-Golgi. In previous studies, we have shown that BODIPY®TR ceramide co-localizes with several different GFP-tagged insect and mammalian N-glycan processing enzymes, including a classic Golgi marker, bovine β4-galactosyltransferase I (11, 12, 89). Thus, the results shown in Fig. 13 indicated that the Tn/H92524GalNAcT-GFP protein is mainly localized in the Golgi apparatus of these baculovirus-infected insect cells. This suggests that the native protein is predominantly localized in the insect cell Golgi apparatus, as well, though it is formally possible that addition of the GFP tag could have altered its intracellular distribution.

DISCUSSION

Generally, the N-glycan structures of insect glycoproteins suggest that insect protein N-glycosylation pathways are truncated versions of the mammalian pathway, with all of the N-glycan trimming reactions, but few of the elongation reactions (reviewed in Refs. 1, 2, 25–27). However, there is growing evidence that this is an overly simplistic view, as insect cells
apparently can process N-glycans more extensively than has been previously recognized. Thus, we are beginning to appreciate that at least some insect cells, under some conditions, have the potential to produce elongated N-glycans with hybrid or complex structures. We believe the N-glycan processing potential of insect cells will be fully appreciated only after insect glycosyltransferase genes have been molecularly cloned and the gene products functionally analyzed. Whereas some published studies have utilized this general approach, most have focused on the processing exoglycosidases (8–17), the core fucosyltransferase, (22), and the early acting glycosyltransferase, N-acetylglucosaminyltransferase I (45). The only published study to date that describes an insect terminal glycosyltransferase gene together with a supporting functional analysis of the gene product is a recent publication on a D. melanogaster sialyltransferase (46). This study provided exciting new evidence that insect cells have the potential to produce complex N-glycans.

The cabbage looper, T. ni, was chosen as the model insect system for this project because a previous study had shown that an established T. ni cell line has low levels of galactosyltransferase and N-acetylgalactosaminyltransferase activities (30). In addition, several studies had shown that T. ni cells could produce recombinant glycoproteins with terminally galactosylated N-glycans when infected with baculovirus expression vectors (32, 35, 38, 39). Thus, it was theoretically possible to isolate a β4-galactosyltransferase family member from this organism.

PCRs with degenerate primers corresponding to conserved regions of β4-galactosyltransferase family members yielded a related DNA fragment, which ultimately led to the isolation of a full-length cDNA from a T. ni library. Bioinformatic analysis of the derived amino acid sequence suggested that this cDNA encoded a membrane glycoprotein with type II topology, which is characteristic of all known Golgi glycosyltransferases (90). Furthermore, the predicted protein had significant similarity to known and predicted β4-galactosyltransferase family members. The closest relatives of the derived T. ni protein were putative β4-galactosyltransferase family members from mos-
quito and fly. The closest relative of known function was a β4-N-acetylgalactosaminyltransferase from *C. elegans* (48). Thus, it appeared that we had molecularly cloned a β4-galactosyltransferase family member from *T. ni*, but it appeared to encode a β4-N-acetylgalactosaminyltransferase rather than a β4-galactosyltransferase.

A baculovirus vector was used to express the full-length *T. ni* cDNA product for biochemical assays of crude enzyme preparations. The results showed that this form of the enzyme could transfer galactose, N-acetylglucosamine, and N-acetylgalactosamine from their respective nucleotide sugar donors to the artificial acceptor substrate, pNP-GlcNAc, *in vitro*, but had a clear preference for N-acetylglactosamine. These results were confirmed by additional *in vitro* assays of a purified affinity-tagged soluble domain of the cDNA product. This form of the enzyme transferred galactose, N-acetylglucosamine, and N-acetylgalactosamine to pNP-GlcNAc *in vitro*. It also transferred all three sugars to glycoprotein and glycolipid acceptors *in vitro*, with a strong preference for N-acetylglactosamine. These results indicated that the *T. ni* cDNA encodes a β4-N-acetylgalactosaminyltransferase with an unusually broad substrate specificity that could potentially function in N-glycoprotein biosynthesis, glycolipid biosynthesis, or both. Because our research interests focus on insect protein glycosylation path-

**Fig. 8.** Glycosyltransferase activity toward a synthetic N-glycan acceptor substrate. Aliquots of the purified soluble, HPC4-tagged *T. ni* cDNA product were incubated with GnGn-PA in the presence of UDP-galactose (B), UDP-N-acetylglucosamine (C), or UDP-N-acetylgalactosamine (D), then the reaction products were analyzed by HPLC with fluorescence detection, as described under “Experimental Procedures.” A, elution profile of the Gn2M acceptor prior to treatment.
ways, the remainder of the present study focused on the potential role of this gene product in N-glycoprotein biosynthesis. However, the broader significance of the results obtained with the glycolipid acceptor is worthy of further discussion. Recent studies have shown that the acquisition of resistance to Cry5B, an invertebrate-specific toxin produced by Bacillus thuringiensis, results from mutations in four C. elegans genes designated bre-2, -3, -4, and -5 (91, 92). Each of these genes encodes known or putative glycosyltransferases. Moreover, bre-4 specifically encodes the C. elegans N-acetylgalactosaminyltransferase that is most closely related to the T. ni enzyme identified in this study. It has been proposed that the enzymes encoded by bre-3, -5, and -4 catalyze a stepwise series of glycosyltransferase reactions culminating with the production of a glycosphingolipid (GalNAcβ4GlcNAcβ3Manβ4Glc-ceramide) required for Cry5B toxicity. The bre-4 product has been implicated in N-glycoprotein (48), but not glycolipid biosynthesis. Therefore, our finding that the T. ni enzyme can transfer N-acetylgalacto-

**Fig. 9.** MALDI-TOF analyses of the N-glycan intermediate and reaction product. A preparative N-acetylgalactosaminyltransferase reaction was performed and the major N-glycans eluting around 20.38 min (A) and 21.35 min (B) in Fig. 8D were recovered from an HPLC run. Each was then desalted and analyzed by MALDI-TOF mass spectroscopy, as described under “Experimental Procedures.” Because of the need to perform scaled up, partial glycosyltransferase reactions to obtain adequate amounts of the putative intermediate eluting around 20.38 min in Fig. 8, the intermediate was contaminated with large amounts of the GnGn-PA acceptor substrate (~1417 in Fig. 9A), and the end product was contaminated with large amounts of the intermediate (~1620 in B).
tosamine to both glycoprotein and glycolipid acceptors in vitro provides at least some evidence to support the idea that the *C. elegans* enzyme might participate in both glycoprotein and glycolipid biosynthesis, as well.

Focusing on its potential role in N-glycoprotein biosynthesis, we found that the *T. ni* enzyme could transfer N-acetylglactosamine, but no detectable N-acetylglucosamine or galactose, to a synthetic bi-antennary N-glycan in vitro. The enzyme produced one or both mono-N-acetylgalactosaminylated intermediates and a di-N-acetylgalactosaminylated product, as observed by HPLC. The masses of both the intermediate(s) and the product were verified by MALDI-TOF and the terminal N-acetylglactosamine residues were shown to be β1,4-linked by exoglycosidase digestions and NMR analysis. These results conclusively demonstrated that the *T. ni* cDNA encodes a β4-N-acetylglactosaminyltransferase, designated Tnβ4GalNAcT, and strongly supported the idea that it functions in insect N-glycoprotein biosynthesis.

This tentative conclusion was further supported by *in vivo* experiments in which Tnβ4GalNAcT was co-expressed with a
reporter N-glycoprotein in baculovirus-infected insect cells. The results showed that Tnβ4GalNAcT transferred N-acetyl-galactosamine, but no detectable galactose or N-acetylglucosamine, to the reporter. Finally, a Tnβ4GalNAcT-GFP fusion protein was localized predominantly in the Golgi apparatus of baculovirus-infected insect cells. Thus, all of the in silico, in vitro, and in vivo data obtained in this study were consistent with the conclusion that the new T. ni cDNA encodes an enzymatically active β4-N-acetylgalactosaminyltransferase which functions in N-glycoprotein biosynthesis. More specifically, Tnβ4GalNAcT yields the N-glycan GalNAcβ4GlcNAc-R, also known as LacDiNAc or LDN. This disaccharide is an increasingly well-recognized alternative to the more common outer chain structure, Galβ4GlcNAc-R, also known as LacNAc, or

**Fig. 11. NMR analysis of the N-glycan reaction product.** A preparative N-acetylgalactosaminyltransferase reaction was performed, and the major N-glycan product, which eluted around 21.35 min in Fig. 8D, was recovered from an HPLC run. The glycan was then desalted, exchanged with D$_2$O, and analyzed by 500-MHz $^1$H NMR spectroscopy, as described under “Experimental Procedures.” A, $^1$H NMR spectrum of the glycan product (500 mHz, 27 °C, D$_2$O). B, resolution-enhanced H-1 and NAc regions. *Italicized numbers* in the spectrum refer to the corresponding residues in the structure.
LN, which can be found in both vertebrate and invertebrate glycoproteins (49). Among the vertebrates, LDN is a common terminal disaccharide of pituitary hormone N-glycans, in which the N-acetylglucosamine residues are 4-O-sulfated and function as recognition markers for hormonal clearance (93, 94). Among the invertebrates, LDN is a common antigenic determinant of parasitic worms (reviewed in Ref. 95) and also has been documented in native T. ni glycoproteins (43). The presence of a β4-N-acetylgalactosaminyltransferase gene in T. ni cells and the antigenicity of worm LDN residues raise important questions about the potential antigenicity of recombinant glycoproteins produced in these insect cells. To date, we are unaware of any structural evidence that LacDiNAc has been found in the N-glycans of any recombinant glycoprotein produced in baculovirus-infected T. ni cells or larvae. In fact, considering that the expression of most or all host cell genes is repressed during baculovirus infection (96), and that β4-N-acetylgalactosaminyltransferase activity is dramatically reduced during baculovirus infection (30), it seems rather unlikely that LacDiNAc will occur in recombinant N-glycoproteins produced using the baculovirus expression system.

It is interesting to consider in retrospect that we actually intended to isolate an insect β4-galactosyltransferase cDNA in this study. It appears that we did not isolate this cDNA, suggesting that T. ni encodes another β4-galactosyltransferase family member, which is responsible for the galactosyltransferase activity observed in cell lysates (30) and the terminal galactose residues observed on some N-glycans produced in these cells (32, 38, 39). Alternatively, it is possible that the cDNA isolated in this study encodes a bifunctional enzyme that has both β4-N-acetylgalactosaminyltransferase and β4-galactosyltransferase activities in vivo. In fact, we detected some galactosyltransferase activity in lysates from insect cells expressing this cDNA product, but it was much lower than the N-acetylgalactosaminyltransferase activity detected in these same lysates. van Die et al. (30) observed the same quantitative relationship between these two endogenous activities in T. ni cell lysates (30). Thus, it is possible that the T. ni cDNA product can also function as a β4-galactosyltransferase activity in vivo, but only at levels below the sensitivity of our lectin blotting assays. It will be interesting to address this possibility in future studies. However, irrespective of whether or not the cDNA isolated in this study encodes a bifunctional enzyme, its isolation, together with functional characterization of the gene product, are significant because the results provide unequivocal, molecular genetic evidence that insect cells have the potential to produce complex N-glycans. The presence or absence of this capability has been a controversial subject in insect glyobiology for the past decade, and the old idea that insect cells cannot produce complex N-glycans has begun to crumble under the growing weight of evidence to the contrary.

**FIG. 12.** In vivo analysis of the role of Tnβ4GalNAcT in N-glucosaminyltransferase activities in insect cells. Sf9 cells were co-infected with recombinant baculoviruses encoding GST-SfManI plus a wild-type baculovirus (Control), a recombinant baculovirus encoding bovine β4-galactosyltransferase I (BovineB4GalT), or a recombinant baculovirus encoding the full-length, untagged T. ni cDNA product (Tnβ4GalNAcT). GST-SfManI was affinity-purified from each source, treated with buffer alone (−) or PNGase-F (+), and then the samples were resolved by SDS-PAGE and analyzed by immunoblotting or lectin blotting, as described under “Experimental Procedures.” A, immunoblots with anti-GST; B, lectin blots with WFA (specific for N-acetylglucosamine); C, lectin blots with RCA (specific for galactose). The positions of molecular mass standards are indicated by their molecular masses in kDa on the left-hand side of the figure.

**Fig. 13.** Intracellular distribution of a Tnβ4GalNAcT-GFP fusion protein. Sf9 cells were infected with a recombinant baculovirus encoding Tnβ4GalNAcT fused to GFP at its C terminus, then the cells were stained with BODIPY®TR ceramide and examined with a confocal laser-scanning microscope, as described under “Experimental Procedures.” A, GFP fluorescence; B, overlay (Merge); C, BODIPY-ceramide fluorescence.
present study, together with the recently published study on a fly sialyltransferase gene coupled with functional characterization of the gene product, should contribute to the demise of this old idea.

Acknowledgments—The fluorescence detector used for the natural substrate assays performed in this study was provided by a faculty mintignant awarded under the auspices of a National Institutes of Health Biomedical Research Infrastructure (BRIN) grant to the University of Wisconsin—Madison (NIH P41 RR014742), and the confocal microscopy and MALDI-TOP analyses were performed in facilities supported by two Center of Biomedical Research Excellence (COBRE) grants to the University of Wisconsin (RR016440 and RR15553). We thank Justin Jones for performing the MALDI-TOP, and we are extremely grateful to Dr. Herman van Halbeek, director of the NMR facility in the College of Chemistry at the University of California–Berkeley for performing the NMR analysis. We also thank Dr. Paul Friedrich of the University of Wisconsin for providing the T. ni cDNA library, Dr. Tadashi Takeki of Shiga University for providing the glycolipid acceptor, Dr. Kwan Kawar of the University of Oklahoma Health Sciences Center for helpful discussions during the course of this work, and HyClone for providing SPX-INSECT and TNM-FH medium.

Note Added in Proof—It should be noted that Narimatsu and coworkers have recently reported the isolation of two human β1,4-N-acetylgalactosaminyltransferases genes (105, 106). Furthermore, they experimentally verified that both gene products were β1,4-N-acetylgalactosaminyltransferases involved in the biosynthesis of LacCNAc. However, the amino acid sequences of these enzymes are not very homologous to those of the C. elegans or T. ni enzymes described in the present study. In addition, the human enzymes contain over twice as many amino acids as the worm and insect enzymes. Finally, the human enzymes were implicated in both N- and O-glycosylation, whereas the T. ni enzyme described in this study was implicated in N-glycan and glycolipid biosynthesis. Thus, the relationship between the enzyme described in this study and the human enzymes described by Narimatsu’s group are unclear at this time.

REFERENCES
1. Marx, L., Altmann, F., Stauchacker, E., and Kubelka, V. (1995) in Glycoconjugates (Montreuil, J., Vliegenthart, J. F. G., and Schachter, H., eds) Vol. 29a, pp. 369–377, Elsevier, Amsterdam
2. Marchal, I., Jarvis, D. L., Cacan, R., and Verbert, A. (2001) Biochim. Biophys. Acta 1519, 154–159
3. Wilson, I. B. H. (2002) Curr. Opin. Struct. Biol. 12, 569–577
4. Jarvis, D. L., Fleming, J. A., Kovacs, G. R., Summers, M. D., and Guarino, L. A. (1990) Bio/Technology 8, 950–955
5. O’Reilly, D. R., Miller, L. K., and Lucwicz, Y. A. (1992) Bacterial Expression Vectors, W.H. Freeman and Company, New York
6. Summers, M. D., and Seidah, G. N. (1988) Agric. Exp. Stn. Bull. No. 1555
7. Jarvis, D. L. (1997) in The Bacteriocides (Miller, L. K., ed.), pp. 389–431, Plenum Press, New York
8. Kawar, Z., Herscovics, A., and Jarvis, D. L. (1997) Glycobiology 7, 433–443
9. Jarvis, D. L., Bohlemy, D. A., Liao, Y. F., Lomax, K. K., Merkle, R. K., Weinkauf, C., and Moremen, K. W. (1997) Glycobiology 7, 113–127
10. Kawar, Z., Romero, P., Herscovic, A., and Jarvis, D. L. (2000) Glycobiology 10, 347–355
11. Kawar, Z., S. van Die, I., Joziasse, D. H., and Jarvis, D. L. (2001) Insect Biochem. Mol. Biol. 31, 289–297
12. Kawar, Z. S., Moremen, K. W., and Jarvis, D. L. (2001) J. Biol. Chem. 276, 16355–16360
13. Francis, B. R., Paquin, L., Weinkauf, C., and Jarvis, D. L. (2002) Glycobiology 12, 369–377
14. Foster, J. M., Yudkin, B., Lockyer, A. E., and Roberts, D. B. (1995) Gene (Amst.) 154, 183–186
15. Rabouille, C., Kuzn, D. A., Lockyer, A., Watson, R., Signerelli, T., Rose, D. F., van den Hevel, M., and Roberts, D. B. (1999) J. Cell Sci. 112, 3319–3330
16. Kerscher, S., Alberti, S., Wuerchpot, H., de Heispen, M., and Schuwe, L. (1995) Dev. Biol. 168, 613–626
17. van den Elzen, J., Kuzn, D. A., and Rose, D. F. (2001) EMBO J. 20, 3008–3017
18. Altmann, F., and Marz, L. (1995) Glycocon. J. 12, 150–155
19. Ren, J., Brethrauer, R. K., and Castellino, F. J. (1995) Biochemistry 34, 2449–2455
20. Ren, J., Castellino, F. J., and Brethrauer, R. K. (1997) Biochem. J. 315, 951–956
21. Davidson, D. J., Brethrauer, R. K., and Castellino, F. J. (1991) Biochemistry 30, 8311–8316
22. Fabini, G., Freiling, A., Altmann, F., and Wilson, I. B. (2001) J. Biol. Chem. 276, 20658–20667
23. Strathclyde, E., Kubelka, V., and Marz, L. (1992) Eur. J. Biochem. 207, 987–993
24. Kubelka, V., Altmann, F., Kornfeld, G., and Marz, L. (1994) Arch. Biochem. Biophys. 308, 149–157
25. Altmann, F., Staudacher, E., Wilson, I. B., and Marz, L. (1999) Glycocon. J. 16, 109–123
26. Jarvis, D. L. (2003) Virology 310, 1–7
27. Tomiya, N., Benetahung, M. J., and Lee, Y. C. (2003) Acc. Chem. Res. 36, 613–620
28. Altmann, F., Kornfeld, G., Dalik, T., Stauchacker, E., and Glash, J. (1993)}
33518  

Cloning and Functional Analysis of Tnβ4GalNAcT

82. Felsenstein, J. (1989) Cladistics 5, 164–166
83. van Halbeek, H. (1994) Methods Enzymol. 230, 132–168
84. Yan, S. B., Chao, Y. B., and van Halbeek, H. (1993) Glycobiology 3, 597–608
85. Hollister, J. R., Grabenhorst, E., Nimtz, M., Conradt, H. O., and Jarvis, D. L. (2002) Biochemistry 41, 15093–15104
86. Hollister, J. R., Conradt, H. O., and Jarvis, D. L. (2003) Glycobiology 13, 487–495
87. Kumar, R., Yang, J., Larsen, R. D., and Stanley, P. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 9948–9952
88. Jarvis, D. L., and Summers, M. D. (1989) Mol. Cell. Biol. 9, 214–223
89. Vadaie, N., Hulinsky, R., and Jarvis, D. L. (2002) Glycobiology 12, 589–597
90. Paulson, J. C., and Colley, K. J. (1989) J. Biol. Chem. 264, 17615–17618
91. Griffiths, J. S., Huffman, D. L., Whitacre, J. L., Barrows, B. D., Marroquin, L. D., Muller, R., Brown, J. R., Hennet, T., Esko, J. D., and Aroian, R. V. (2003) J. Biol. Chem. 278, 45594–45602
92. Griffiths, J. S., Whitacre, J. L., Stevens, D. E., and Aroian, R. V. (2001) Science 293, 869–874
93. Fiete, D. J., Beranek, M. C., and Baenziger, J. U. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 2089–2093
94. Smith, P. L., Bousfield, G. R., Kumar, S., Fiete, D., and Baenziger, J. U. (1993) J. Biol. Chem. 268, 795–802
95. Cummings, R. D., and Nyame, A. (1999) Biochem. Biophys. Acta 1455, 363–374
96. Nohiron, I., O’Heilly, D. R., and Olszewski, J. A. (2003) J. Gen. Virol. 84, 3029–3039
97. Massi, R. A., Appert, H. E., and Fukuda, M. N. (1988) Biochem. Biophys. Res. Commun. 157, 657–663
98. Schwientek, T., Almeida, R., Levery, S. B., Holmes, E. H., Bennett, E., and Clausen, H. (1998) J. Biol. Chem. 273, 29331–29340
99. Almeida, R., Levery, S. B., Mandel, U., Kresse, H., Schwientek, T., Bennett, E. P., and Clausen, H. (1999) J. Biol. Chem. 274, 26165–26171
100. Okajima, T., Yoshida, K., Kondo, T., and Furukawa, K. (1999) J. Biol. Chem. 274, 22015–22021
101. Adams, M. D., et al. (2000) Science 287, 2185–2195
102. Nakamura, Y., Haines, N., Chen, J., Okajima, T., Furukawa, K., Urano, T., Stanely, P., Irvine, K. D., and Furukawa, K. (2002) J. Biol. Chem. 277, 46280–46288
103. Takemae, H., Ueda, R., Okubo, R., Nakato, H., Izumi, S., Saigo, K., and Nishihara, S. (2003) J. Biol. Chem. 278, 15571–15578
104. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nucleic Acids Res. 22, 4673–4680
105. Sato, T., Gotoh, M., Kiyohara, K., Kameyama, A., Kubota, T., Kikuchi, N., Ishizuka, Y., Iwassaki, H., Totagayachi, A., Kudo, T., Ohkura, T., Nakasho, H., and Narimatsu, H. (2003) J. Biol. Chem. 278, 47534–47544
106. Gotoh, M., Sato, T., Kiyohara, K., Kameyama, A., Kikuchi, N., Kwon, Y. D., Ishizuka, Y., Iwai, T., Nakasho, H., and Narimatsu, H. (2004) FEBS Lett. 562, 134–140
Molecular Cloning and Functional Characterization of a Lepidopteran Insect β4-N-Acetylgalactosaminyltransferase with Broad Substrate Specificity, a Functional Role in Glycoprotein Biosynthesis, and a Potential Functional Role in Glycolipid Biosynthesis

Nadia Vadaie and Donald L. Jarvis

J. Biol. Chem. 2004, 279:33501-33518.
doi: 10.1074/jbc.M404925200 originally published online June 1, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M404925200

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 100 references, 32 of which can be accessed free at http://www.jbc.org/content/279/32/33501.full.html#ref-list-1