Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Kristoffer Leon 04/15/15

Cloning, expression and characterization of a β1,4-GalNAcTransferase from *Schistosoma mansoni*.

by

Kristoffer E. Leon

Richard D. Cummings, Ph.D.
Department of Biochemistry
Adviser

Department of Chemistry

Richard D. Cummings, Ph.D.
Department of Biochemistry
Adviser

José Soria, Ph.D.
Department of Chemistry
Committee Member

Emily Weinert, Ph.D.
Department of Chemistry
Committee Member

2015

Cloning, expression and characterization of a β1,4-GalNAcTransferase from *Schistosoma mansoni*.

By

Kristoffer E. Leon

Richard D. Cummings, Ph.D.
Department of Biochemistry
Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Chemistry

2015

Abstract

Cloning, expression and characterization of a β1,4-GalNAcTransferase from *Schistosoma mansoni*.

By Kristoffer E. Leon

Schistosomiasis is an insidious parasitic helminth infection that affects millions of people worldwide. Infection is marked by a number of clinical manifestations, ranging from dermatitis to chronic inflammation of the liver and spleen. *Schistosoma mansoni* is one of the helminth species responsible for schistosomiasis, and has been previously studied for its interactions with the human immune system and its carbohydrate (glycan) biochemistry. The entire genome has recently been published online and partially annotated. Computational tools were used to select a gene coding for a β1,4-N-acetylgalactosaminyltransferase (smβ1,4-GalNAcT), a class of enzymes responsible for synthesizing LacDiNac (LDN). LDN plays an important role in the immune response to *S. mansoni* and is a parent glycan that is further modified by the addition of fucose(s) to generate other antigenic carbohydrates. The gene was cloned into a mammalian expression vector, expressed in a mammalian cell line grown in suspension, purified and then used in activity assays to prove the computationally determined gene was a smβ1,4-GalNAcT. Activity assays were also used to determine metal cofactor necessity, and the specificity for the nucleotide sugar donor UDP-GalNAc. The enzyme itself is an important member of *S. mansoni* glycosylation pathways and immunology, and could potentially represent a target for pharmaceutical development. Lastly, the smβ1,4-GalNAcT provides a link between the glycome and genome. Glycomics is becoming more important, especially in understanding host-pathogen interactions, and knowing the enzymes necessary for glycan synthesis is important in understanding the complex and fascinating glycobiology of *S. mansoni*.

Cloning, expression and characterization of a β1,4-GalNAcTransferase from *Schistosoma mansoni*.

By

Kristoffer E. Leon

Richard D. Cummings, Ph.D.
Department of Biochemistry
Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Chemistry

2015

Acknowledgements

I would like to thank everyone has helped me with this project and my undergraduate career. Megan Mickum and Nina Salinger-Prasanphanich have been working with me for the past two years and have been an integral part of my training towards a career in the sciences. Both have provided an incredible amount of assistance, training and advice while I have been working in the Cummings laboratory. I would especially like to thank Megan for all of her hard work in helping to develop this project and for taking me on as a student, since she helped and taught along every step of the way.

Dr. Cummings has also been an integral part of my career. I appreciate being welcomed into the lab with open arms and an opportunity to perform scientific research on a fascinating helminth. His advising has also helped to shape my career path, and the training in the glycosciences is invaluable.

Dr. Soria was one of my first professors at Emory University. He challenged me like I have never been challenged before, and he helped to convince me that I truly love both chemistry and science. The opportunity to mentor students underneath him has been an amazing opportunity while at Emory.

Dr. Weinert taught one of my favorite courses at Emory University. I was passionate about biochemistry and organic chemistry, and she helped to show how I could meld those two disciplines. Her course helped to develop my proposal for my thesis, as well as the motivation to complete the project.

Table of Contents

Introduction

Schistosomiasis Epidemiology

Schistosomiasis is among the most widespread parasitic infections in the world. Currently, there are over 200 million people at risk of infection with one of these disease causing helminths: *Schistosoma mansoni, Schistosoma haematobium* or *Schistosoma japonicum. S. mansoni* is a widespread parasite found in the Caribbean, the Middle East, South America and Africa (*1*). Furthermore, the parasite is found in several freshwater travel destinations, such as the Lake Malawi, Lake Kariba, and the Zambezi River, and presents as a threat to locals, travelers and livestock (*2*). The impact that *S. mansoni* has on these regions is significant. However, schistosomiasis has a high disease burden outside of mortality, especially on development in young patients and pulmonary and neurological complications from long-term infection (*3*). Furthermore, praziquantel is the only drug currently being used to treat schistosomiasis, which brings about fears of resistance (*4-6*). Furthermore, praziquantel has a variable cure rate, largely dependent on the life stage of the worms and cannot be used for prophylaxis (*6*). Schistosomiasis falls under the definition of a poverty trap, which enforces the low economic status of infected individuals by reducing physical and mental health (*7*). The economic burden caused by chronic schistosomiasis is difficult to estimate, but these serious morbidities cause reduced productivity, maintaining the poverty of those infected with schistosomiasis. *Schistosoma mansoni* infects 54 millions people worldwide, and over 300 million people at risk of infection (*8*). Schistosomiasis is a global

threat and research to understand the disease is necessary to make the steps towards assisting those affected.

Schistosoma mansoni life cycle

Understanding the life cycle of this complex parasite is important for the development of vaccines and therapeutics. *Schistosoma sp*. have a two part life cycle that require both a snail and mammal host. *S. mansoni* infect *Biomphalaria glabrata* snails*. (9)*. These aquatic snails shed cercariae, which are the infectious life stage for mammals. The cercariae are able to enter the human body through mechanical penetration and enzymatic proteolysis (*10-13*). Upon successful penetration of the tissue, the tail is lost, and the cercariae transforms into the juvenile stage of the adult worm known as the schistosomula. The schistosomula travels through the lung vasculature to the final destination of the hepatic and intestinal vasculature (*12, 14*). Over time, the schistosomula mature into an adult form capable of sexual reproduction. Adult worms feed on red blood cells to obtain iron. Adult male and female worms pair in the vasculature and reproduce. Disease presents at the reproductive stage as several hundred eggs are produced daily, and must travel from the circulatory system to the intestinal tract (*15*). The eggs are capable of traveling through several epithelial layers in order to accomplish this feat (*16*). However, not all eggs are successful and travel to other portions of the body, such as the spleen and liver, and cause inflammation and disease (*16, 17*). After successful passage into the intestinal tract, the eggs exit the human host with the fecal matter (*16, 18*). When the fecal matter makes contact with a body of water, the eggs will open, releasing miracidiae, which are capable of infecting snails. Upon infection of

the mollusk host, the miracidiae will develop into the cercariae and continue the life cycle (*16, 18, 19*). While the intestinal tract is important in *S. mansoni* and *S. japonicum* infections, *S. haematobium* instead utilizes the urinary tract for exit from the human host (*18*). The life cycle is complex and presents a multitude of opportunities to learn more about helminth biology. While much work has been done on understanding the life cycle, the knowledge is superficial and there is little known about the molecular determinants of each stage.

Schistosoma mansoni clinical manifestation

Clinical manifestation is directly linked to the different life stages of the disease. The epidemiology of *S. mansoni* is especially relevant to the clinical aspect of the disease due to the destructive and insidious nature of the parasitic infection. Schistosomiasis presents with a variety of symptoms. Initial infection can be marked by cercarial dermatitis or rash, fever, heptaosplenomegaly, fatigue, myalgia, malaise, cough and eosinophilia. In non-endemic persons this disease stage, termed Katyama fever, is more severe and can last anywhere between two to 10 weeks (*19-21*). Chronic pathology is markedly different than acute schistosomiasis, especially since the worms have reached adulthood and egg laying has begun. The eggs induce a robust immune response (*2, 20*). Common symptoms include diarrhea, inflammation, pseudopolyposis, microulcerations and superficial bleeding. More severe versions of chronic *S. mansoni* infection include hepatic, pulmonary, genital and neurological complications caused by either the adult worm or the eggs (*2, 19*). Chronic schistosomiasis has also been linked with anemia, undernutrition, growth

stunting, host immunity, and fertility issues (*22*). Furthermore, there is little to no evidence of the body's ability to clear adult worms (*23*).

Immune response to Schistosomiasis

The immune response to *S. mansoni* is important to the disease pathogenesis. Schistosomiasis is chronic because *S. mansoni* has the ability to modulate the immune system so that the body is unable to clear the adult worm (*16*). Schistosomiasis differentiates between acute and chronic disease, which is due to the immune response to the different life stages. Katayama fever, which is associated with the maturation of the schistosomula to an adult, is seen as a T Helper 1 (T_H1) response to the parasite (*16, 24*). However, chronic disease is associated with a T Helper 2 (T_H2). While a T_H2 response is helpful in dampening the inflammatory response, it allows the adult worm to escape destruction. Furthermore, constant, long-term expression of cytokines associated with a T_H2 eventually results in other diseases associated with long-term inflammation. The shift towards a T_H2 response is commonly seen in other parasitic helminth diseases as well (*25, 26*).

Carbohydrate Immunomodulation

Glycans, or oligosaccharides, are complex carbohydrates linked to proteins (glycoproteins) and lipids (glycolipids) and are the predominant immune targets in *S. mansoni* infection. Previously, it was believed that *S. mansoni* was able to maintain a chronic infection and avoid clearance by glycan mimicry (*27, 28*). Essentially, it

was thought that the helminth created and secreted glycans similar to the host to avoid detection. However, more recent research revised that theory to that of "glycan gimmickry", which is the current working hypothesis (*29*). Carbohydrates are bioactive, and in the case of *S. mansoni*, the glycans directly interact with the immune system and change the nature of the immune response from T_H1 to T_H2 (*16, 30*). The eggs of *S. mansoni* are particularly important in the modulation of the immune system by Asn-linked oligosaccharides (*N*-glycans) and glycolipid glycans (*30, 31*).

Schistosoma mansoni and LDN

While there are many unique glycans produced by *S. mansoni,* the LacDiNAc (LDN) family is of particular interest due to the antigenic nature of its members. LDN is a parent structure that is further modified, particularly by the addition of one or more fucose residues, to create antigens important in immunomodulation (*32-34*). The LDN antigen also generates an antibody response, as well as initiating granuloma formation (*35-37*). LDN modified by fucose addition(s) generate significant immune responses. Several of these structures are shown in **Figure 1**. Responses to these antigens include significant antibody generation and modulation of the immune system (*38-40*).

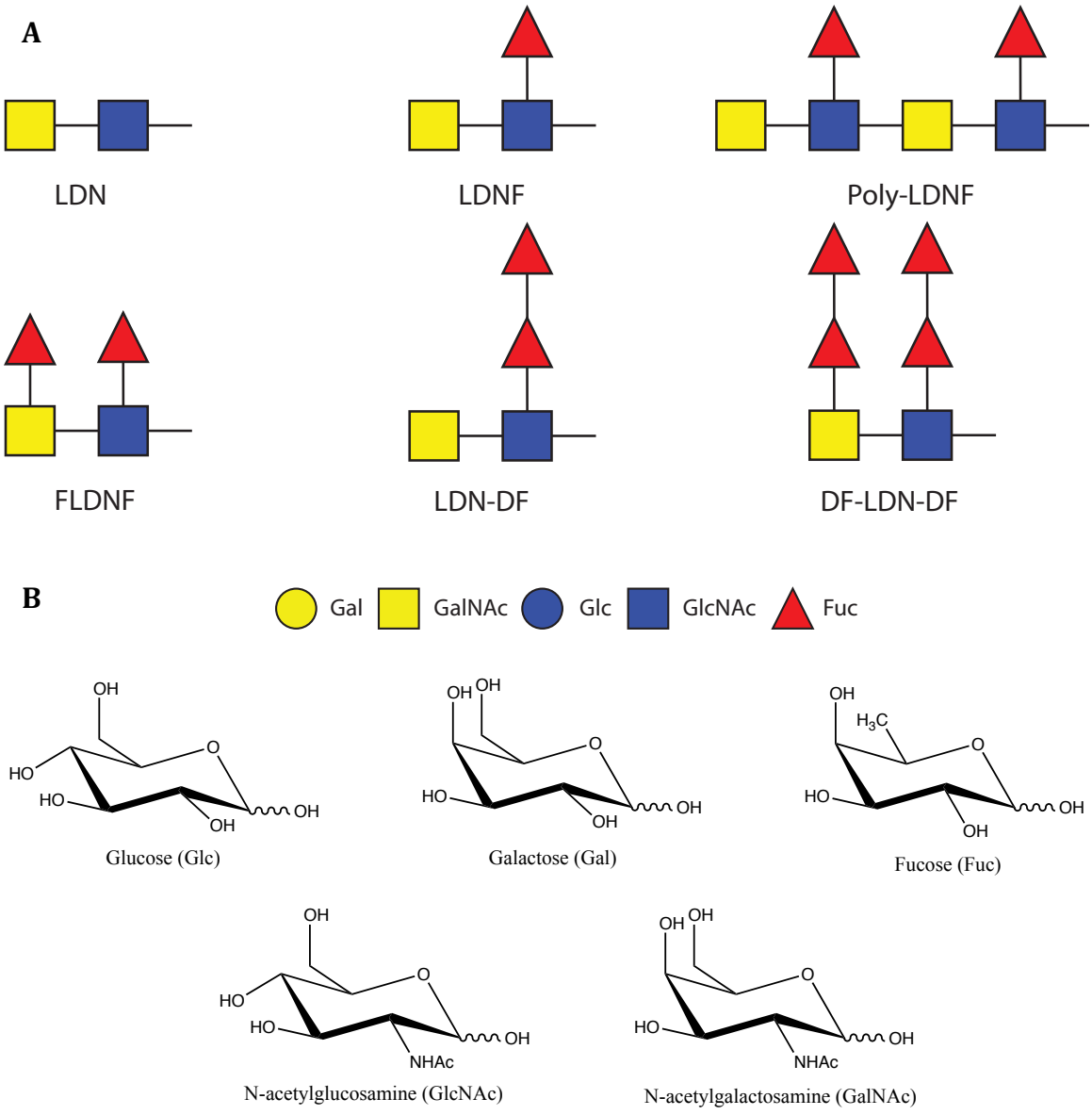

Figure 1: A) The above glycan structures are known antigens of the human immune system and are produced by *S. mansoni*. B) The chemical structures of each sugar.

The glycogenome and glycosyltransferases

The glycans produced by *S. mansoni* are extraordinarily important in the development of the immune response to the parasite and the clinical manifestation of schistosomiasis. Therefore, understanding the molecular processes necessary in synthesizing these antigens is extremely important. The genome for *S. mansoni* has

recently been published, partially annotated, and augmented with RNA-Seq data (*41, 42*). Several of the transcripts have been identified as glycosyltransferases, which are enzymes that synthesize glycans. A brief schema of glycosyltransferase function is shown in **Figure 2.** Three of these glycosyltransferases were further identified as β1-4 N-acetylgalactosaminyltransferases (β1-4GalNAcT), a family of enzymes that are responsible for the synthesis of LDN (*43, 44*). The ortholog from *C. elegans* has previously been expressed, cloned and shown to be functional in generating LDN (*45*). The β1-4GalNAcT from *S. mansoni* could be a potential target for therapeutics.

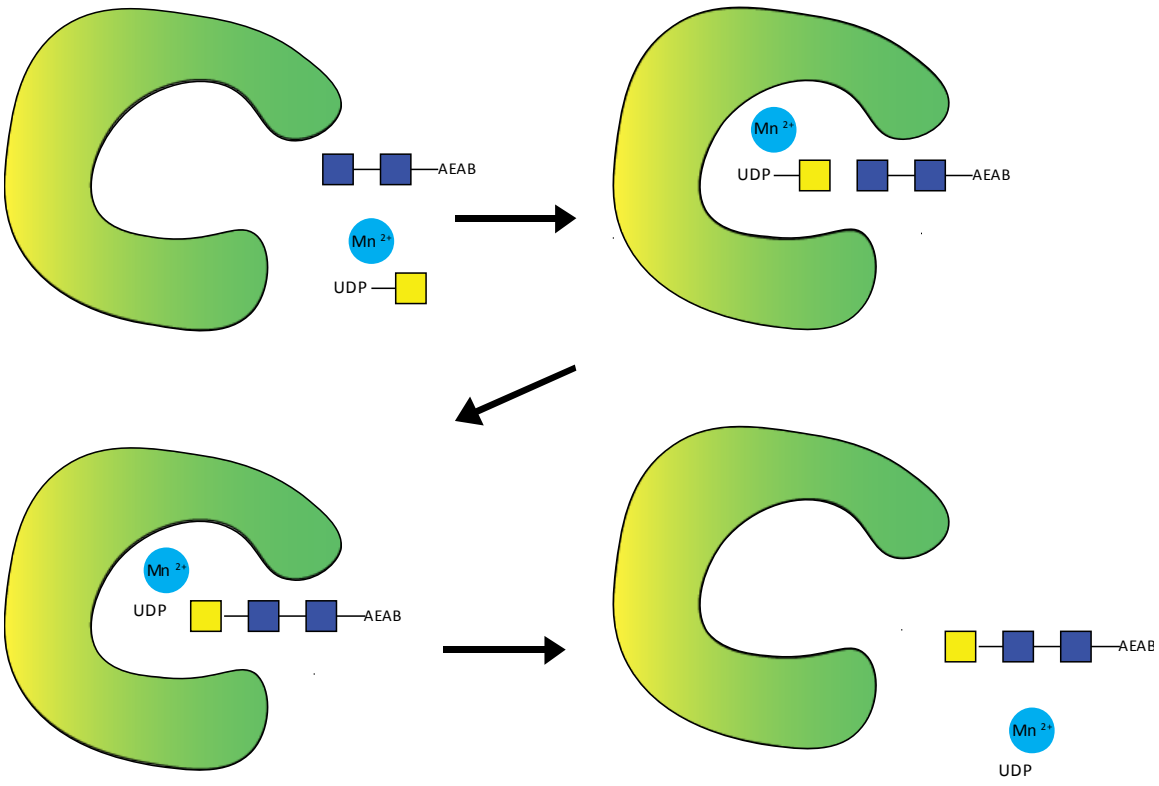

Figure 2: The glycosyltransferase (yellow-green) binds Mn^{2+}, the UDP-GalNAc donor (yellow), and the chitobiose acceptor (blue). The enzyme catalyzes the addition of the GalNAc to the chitobiose acceptor to generate the product through the cleavage of the UDP. The Mn^{2+} ion associates with the phosphate of the nucleotide to stabilize binding to the glycosyltransferase.

The following thesis describes the process of cloning, expressing and characterizing a *S. mansoni* β1-4GalNAcTransferase. First, potential β1-4GalNAcT transcripts were identified based on the annotated genomic data and homology with ceβ1-4GalNAcT. Then, computational analysis of RNA-Seq data for transcript expression is reported. A single gene was targeted for cloning and expression, which was performed using a human expression plasmid containing a GFP reporter tag. Following the cloning, transfection of HEK Freestyle cells grown in suspension with expression of the protein is reported. After purification from cellular lysate, the β1-4GalNAcT was assayed with a variety of acceptor and donor pairs, as well as at different temperatures and with varying co-factors. The conclusion arrives at the successful characterization of a novel glycosyltransferase from a parasitic helminth that causes morbidity and mortality worldwide.

Materials and Methods

All chemicals were purchased from Sigma-Aldrich, restriction enzymes were purchased from New England BioLabs, and DNA extraction and clean-up kits from Qiagen unless noted otherwise.

Selection of a putative β1,4GalNAcT

The entire genome of *S. mansoni* has recently been published, with partial annotations included for several genes. Many of the genes have been predicted based off of computational techniques. Several keywords (GalNAc, N-acetylgalactosamine, etc.) were used to search for GalNAcTransferases within the

GeneDB database for schistosomiasis (http://www.genedb.org/Homepage/Smansoni). Seven putative GalNAcTransferases were identified. The coding domain sequences were translated *in silico* and aligned against the previously identified β1-4GalNAcTransfease from *C. elegans*. Translations, alignments and phylogenetic trees were generated using Geneious V4.7.6. For the alignments, the cost matrix was 70%, with a gap open penalty of 12, a gap extension penalty of 3 and a global alignment type with free end gaps. For the phylogenetic tree, the Jukes-Cantor was used to determine genetic distance, with a neighbor joining build. The entire sequence Smp 021370, termed smβ1-4GalNAcT, was synthesized by Genewiz with flanking EcoRI and BamHI restriction sites in a pUC57 vector.

Cloning of smβ1-4GalNAcT

The gene was excised from the pUC57 vector using ECORI-HF and BAMHI-HF for an hour at 37°C to ensure complete digestion. smβ1-4GalNAcT was separated from the pUC57 backbone by agarose gel and excised and extracted using a QIAquick gel extraction kit. smβ1-4GalNAcT was ligated to a similarly digested expression vector for mammalian cells, termed pGen2, overnight at 16°C. This plasmid construct is now referred to as smβ1-4GalNAcT -pGEn2 and its protein product smβ1-4GalNAcT-GFP. GalNacT-pGEn2 was transformed into E. coli by heat shock and then plated on agar plates with ampicillin at 100ug/ml. Single colonies were used to inoculate culture for plasmid extraction using a QIAquick Spin Mini-prep kit. After extracting the plasmids from the bacteria, the plasmids were digested

with BamHI and EcoRI, ran on an 1% agarose gel for the presence of the smβ1-4GalNAcT gene segment. The plasmids were also sent for Sanger sequencing (Genewiz) to ensure that Smp 021370 was properly inserted. The plasmids were prepared by using a Qiagen Maxiprep kit to obtain DNA concentrations appropriate for transfection of mammalian cells.

Expression of smβ1-4GalNAcT-GFP in HEK-Freestyle Cells

HEK 293 Freestyle cells were grown to a density of $1.8x10^6$ cells/mL in suspension on a platform shaker in a humidified 37°C incubator. Polyethylimine (PEI, Polyscience, Inc.) was used as the transfection agent. A PEI-DNA complex at 1:3 was incubated for 30 minutes and then added to the culture for a final concentration of 3μg/mL of DNA and 9μg/mL of PEI. Twenty-four hours after the PEI-DNA complex was added, the media volume was doubled and valproic acid was added to a final concentration of 2.2 mM. Every day, 200 μL aliquots of cells were frozen for western blot to monitor the expression of the protein. Images were also taken using a fluorescent microscope (Olympus IX51, 10x zoom) to visually monitor GFP fluorescence. At 120 hours, the cells were harvested.

Purification of smβ1-4GalNAcT-GFP

The HEK cell pellet resuspended in Ni-NTA wash buffer (50 mM NaH_2PO_4, 300 mM, 20 mM imidazole, pH 8), protein inhibitor (cOmplete, EDTA-free, Roche) and 0.25% Triton X-100. The cells were then sonicated to disrupt the cell membranes and generate a lysate. After sonication, the cell lysate was spun down

for 5 minutes at 3220xg to clarify the lysate and remove the insoluble materials. The lysate was run over a 2 ml column of Ni-NTA agarose (Qiagen) according to manufacturer's recommendations. GFP positive elution fractions were collected and dialyzed overnight into 45mM HEPES, 300 mM NaCl and 50 mM imidazole using dialysis cassettes with a 10kDa MWCO. The dialyzed solution was then concentrated using a 10kDa MWCO centrifugal filter unit (Millipore) and then stored at 4°C for up to one month. Protein concentrations were approximated using a BCA protein assay kit (Pierce).

Western Blots and Silver Stain

Western blots were performed by adding NuPage loading dye to 1X concentration. The mixture was then boiled for 15 minutes, spun down for 1 minute at 10000xG and then added to a 4-20% SDS-PAGE gradient gel (BioRad). For western blots, the gel was transferred to a nitrocellulose membrane and blocked overnight at 4°C in 5% BSA in TBST (20mM Tris, 300 mM NaCl, pH 7.2, 0.1% Tween-20). The ladder used was the Spectra Multicolor Broad Range Protein ladder (Thermo Scientific). After washing, the anti-GFP (GT859, GeneTex) antibody was incubated for one hour at a 1:10,000 dilution. The secondary, anti-mouse IgG (074-1802, KPL) was incubated at 1:5,000 dilution for 45 minutes and then washed with TBST. Post secondary incubation and washing, SuperSignal West Pico Chemiluminescent substrate (Thermo Scientific) was added to the membrane and then exposed.

Silver stains were run as previously described in reference (*46*).

EndoH (NEB) and PNGaseF (NEB) digestion were performed following the manufacturer's protocol.

Enzyme Assays

Enzymatic reactions were set up in 25 μL quantities. The reactions included the purified protein (1-5 μg), $MnCl_2$ (10 mM), sodium bicarbonate buffer (50 mM), shrimp intestine phosphatase, a donor nucleotide sugar (1mM) and an acceptor AEAB-linked sugar (40μM)(*47*). AEAB is a fluorescent linker. EDTA was used at a final concentration of 10mM. Reactions were run for varying times, typically longer than 24 hours. The reactions were performed at both room temperature and at 37°C while rotating. The different reaction conditions are shown in **Tables 1** and **2**.

The reaction was cleaned via porous graphitized carbon solid phase extraction as previously described in reference (*48*) (Grace Davison Discovery Science), vacuum concentrated (Labconco Centrivap Concentrator) to remove the organic solvent and then lyophilized overnight. The lyophilized material was then re-suspended in 100 μL and then analyzed using HPLC and MALDI.

Table 1: Endpoint Enzyme Assays

Trial	Acceptor	Donor	Ion	Temperature	Duration
1	UDP-GalNAc	Triose	$MnCl_2$	Room Temp.	5 days
2	UDP-GalNAc	Triose	$MnCl_2$	37°C	5 Days
3	UDP-GalNAc	Chitobiose	$MnCl_2$	37°C	5 days
4	UDP-Gal	Triose	$MnCl_2$	37°C	5 days
5	UDP-GalNAc	Triose	EDTA	37°C	93 hours

Table 2: Timepoint Enzyme Assays

Trial	Acceptor	Donor	Ion	Temperature	Duration
1	UDP-GalNAc	Triose	$MnCl_2$	37°C	24 hours
2	UDP-GalNAc	Triose	$MnCl_2$	37°C	48 hours
3	UDP-GalNAc	Triose	$MnCl_2$	37°C	93 hours
4	UDP-GalNAc	Chitobiose	$MnCl_2$	37°C	24 hours
5	UDP-GalNAc	Chitobiose	$MnCl_2$	37°C	48 hours
6	UDP-GalNAc	Chitobiose	$MnCl_2$	37°C	93 hours

HPLC

A Shimadzu HPLC CBM-20A system coupled with a fluorescence detector RF-10AXL was used for HPLC analysis and separation of glycan-AEABs. The glycans were purified by reverse phase HPLC on a porous graphitized carbon column (150 × 4.6 mm, Thermo Scientific) with a gradient of 15–45% acetonitrile (0.1% trifluoroacetic acid) in 40 min. The glycans were monitored by fluorescence at 420 nm.

MALDI-TOF Mass Spectrometry Analysis of Glycans

After HPLC purification of the reactions and the collection of peaks, the carbohydrates were peracetylated and then permethylated as previously described by references (*49*) and (*50*), respectively. The carbohydrates were then analyzed by a matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectrometer. MALDI plates were spotted with 0.5 μL DHB matrix (10 mg/mL DHB, 0.1% Trifluoroacetic Acid) and 0.5 μL of sample (50% methanol).

Results

Selection of a β1-4GalNAcT

Seven genes are annotated in the *S. mansoni* genome database as probable GalNAcTs (**Table 3**). The coding domain of each sequence was extracted from GeneDB, and then translated into a protein sequence. The protein sequences were then aligned to a β1-4GalNAcT from the nematode *C. elegans* (ceGalNAcT) (**Figure 3**). The database is incomplete, and many of the transcripts found appeared to be incomplete and lacking certain features, such as a transmembrane domain.

Table 3: Gene IDs of the GalNAcTs annotated in *S. mansoni* gene database

Putative N-acetylgalactosamine transferases (GalNAcTs)	Gene ID
1	Smp 057620
2	Smp 159490
3	Smp 005500
4	Smp 139230
5	Smp 211240
6	Smp 021370
7	Smp 047240

B

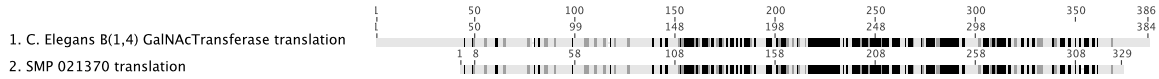

Figure 3: Out of the seven genes, SMP 021370 was found to have both a high level of homology and the most complete sequence at the protein level. The black regions represent identical amino acid residues, dark grey denotes similar residues and light grey are completely different residues. A) The alignment of the seven GalNAcTs using the ceGalNAcT as the reference sequence. B) ceGalNAcT and SMP 021370 aligned against each other to show which regions share homology.

β1-4GalNAcTs require several motifs in their catalytic domain in order to be functional. These include conserved domains that bind nucleotide sugars and often metal ion co-factors. As seen in **Figure 3**, there are regions that share the complete

homology between the two protein sequences, which include binding motifs for a donor-binding site, and a metal binding site as predicted by NCBI's conserved domain database. However, there was no transmembrane domain in the sequence predicted by the TMHMM Server v2.0. Furthermore, Smp 021370 has an isoleucine residue at the same position as ceβ1-4GalNAcT, corresponding to the Tyr-289 in bovine β1-4Gal-T1, which has been shown to have a role in increasing the preference of UDP-GalNAc over UDP-Gal (*51, 52*). In **Figure 4,** a phylogenetic tree was generated using the protein sequences of the seven smGalNAcTransferases, ceGalNAcT and human B4GALNT. Human B4GALNT is responsible for transferring UDP-GalNAc to a GlcNAc acceptor to generate the LDN motif (*53*). In **Figure 4A**, the glycosyltransferases separate into different clades, with Smp 005500, Smp 057620, Smp 159490 and Smp 129230 forming their own monophyletic group. As seen **Figure 4B**, the human B4GALNT is most closely related to SMP 021370. Data from the RNA sequencing was used to generate a plot showing the expression of each of the putative β1-4GalNAcT transcripts in four different points of the life cycles (**Figure 5**)(*42*). The reads per kilobases per million of mapped reads (RPKM) is a normalized value based on the depth of the sequencing that helps to predict expression of a gene in a sample (*54*). The addition of this information assisted in the decision to express and characterize Smp 021370.

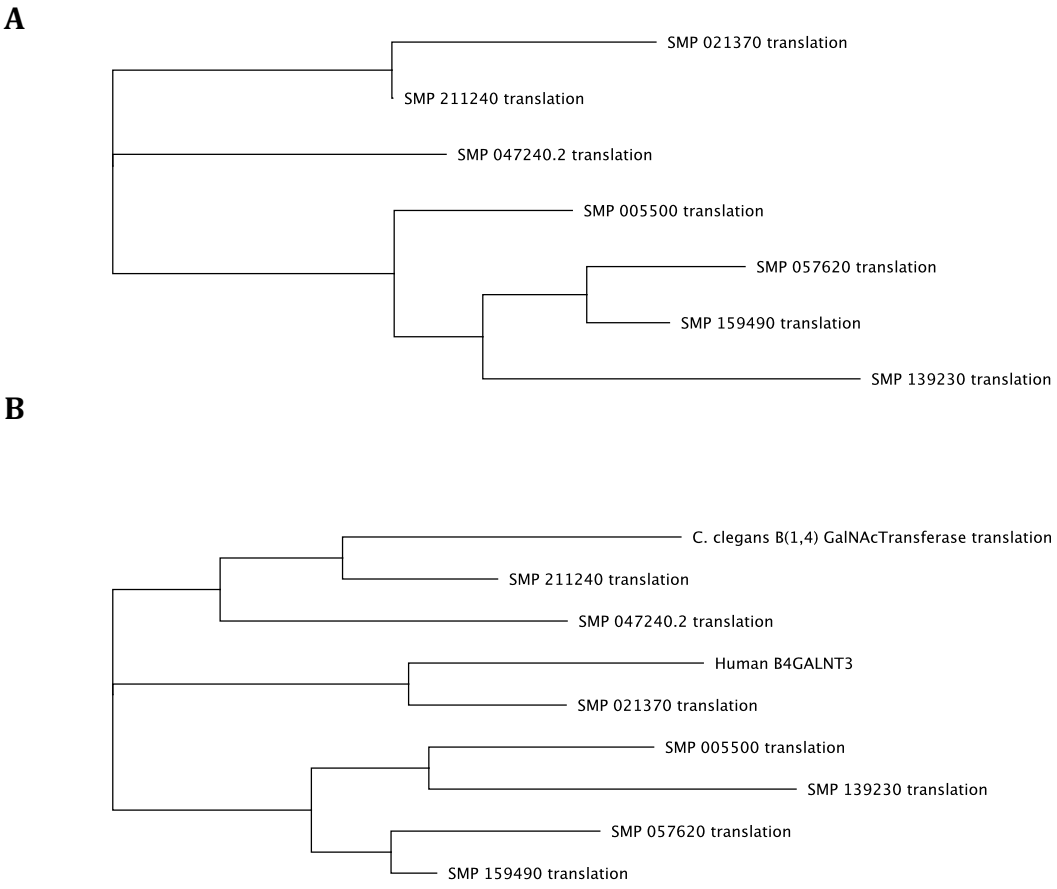

Figure 4: A) Phylogenetic tree based on amino acid sequence of the seven putative GalNAcTs B) Phylogenetic tree based on amino acid sequence with amino acid sequences of ceGalNAcT and hGALNT3 included in analysis. The trees are unrooted.

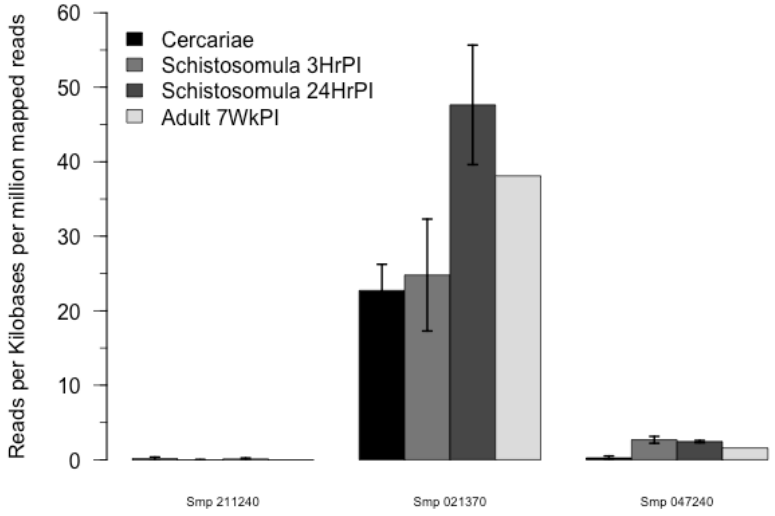

Figure 5: Smp 021370 is the best-represented transcript out of all of the potential GalNAcTs. The error bars are standard deviations.

Based on the presence of UDP-GalNAc and Mn^{2+} binding motifs, homology with ceGalNAcT, phylogenetic relationship to hGALNT3 and the levels of RNA expression, the gene transcript Smp 021370 was synthesized. Smp 021370 will be referred to as smGalNAcT.

Synthesis and Cloning of smGalNAcT

SMP 021370 was synthesized by GeneTec and cloned into a pUC57 vector. The gene was excised by restrictions enzymes, and then ligated into a pGEN2 vector (smGalNacT-pGen2) (**Figure 6**). The pGEN2 vector was designed to generate soluble, secreted GFP fusion enzyme, which can be transiently expressed in mammalian cell lines and purified via Ni-NTA affinity chromatography. The recombinant protein product will be called smβ1-4GalNAcT.

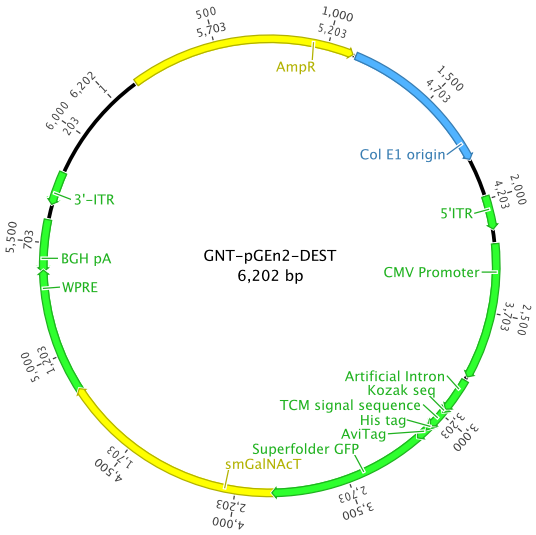

Figure 6A: Vector map of the smGalNAcT-pGEn2 vector. The yellow annotation on the bottom of the map labeled GalNAcT is where the sequence for the gene is, which is adjacent to the superfolder GFP. The mammalian expression vector produces an N-Terminal GFP. The Kozak sequence, TCM signal sequence, His tag, AviTag, Superfolder GFP and smGalNAcT are all transcribed into mRNA and then translated.

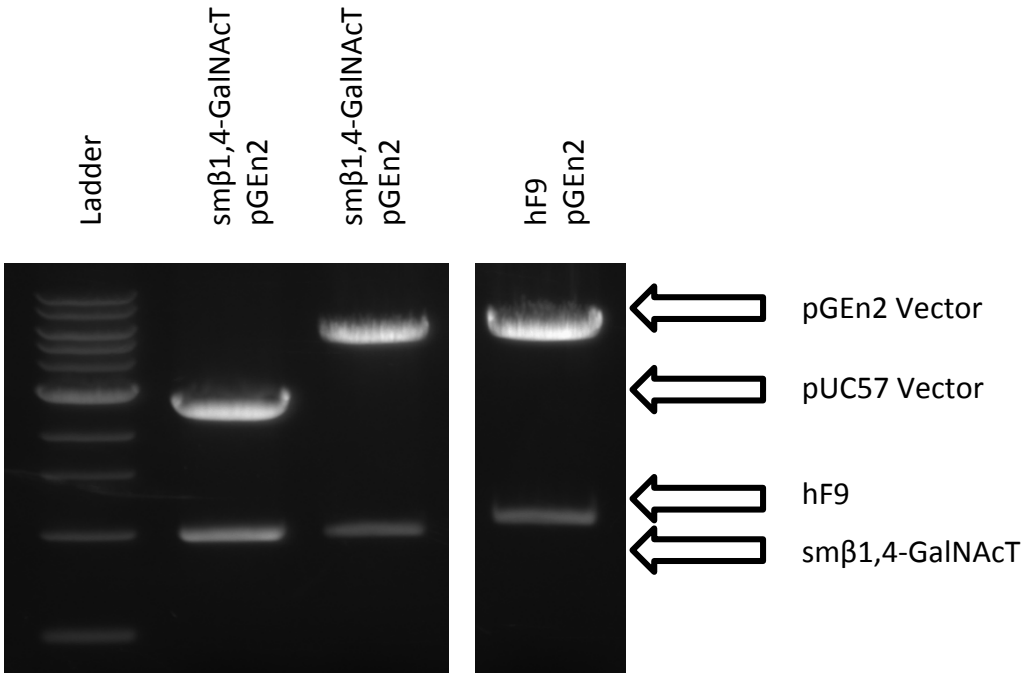

Figure 6B: Restriction digests of the synthetic gene smβ1-4GalNAcT in the pUC57 vector and after ligation into the pGen2 expression vector. The smβ1-4GalNAcT has a length of 987 bases, the gene hF9 (Human Fucosyltransferase 9 or Fut9) includes the attB1 primer site and a TEV cleavage site for a total length of 1061 bases, the pUC57 vector is 2710 bases and the pGEn2 vector has a length of 5215 bases. The ladder is 1 kb DNA ladder.

Sanger sequencing also was used to confirm the insertion of smβ1-4GalNAcT into the pGEN2-DEST vector. After successful plasmid generation, *E. coli* top10 cells were transformed and the smβ1-4GalNAcT -pGen2 isolated using a Qiagen Maxi Kit.

Expression of smβ1-4GalNAcT into HEK-Freestyle cells

Human Embryonic Kidney 293-Freestyle (HEK293-F) cells (a derivative of HEK 293 cells designed for suspension growth) were transfected using the smβ1-4GalNAcT -pGEn2 and PEI Transfection efficacy was monitored via flow cytometry and fluorescent microscopy of GFP (Figure 7).

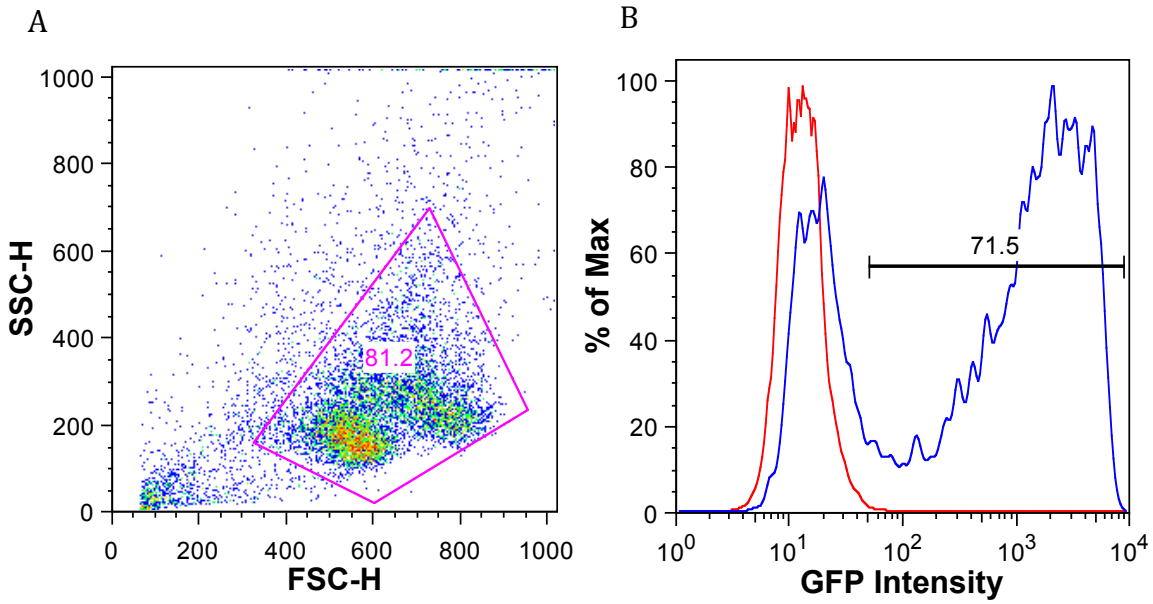

Figure 7: A) Gating of living cells on flow using untransfected HEK cells B) An overlay of untransfected cells (red) and cells transfected with smβ1-4GalNAcT - pGEn2 (blue).

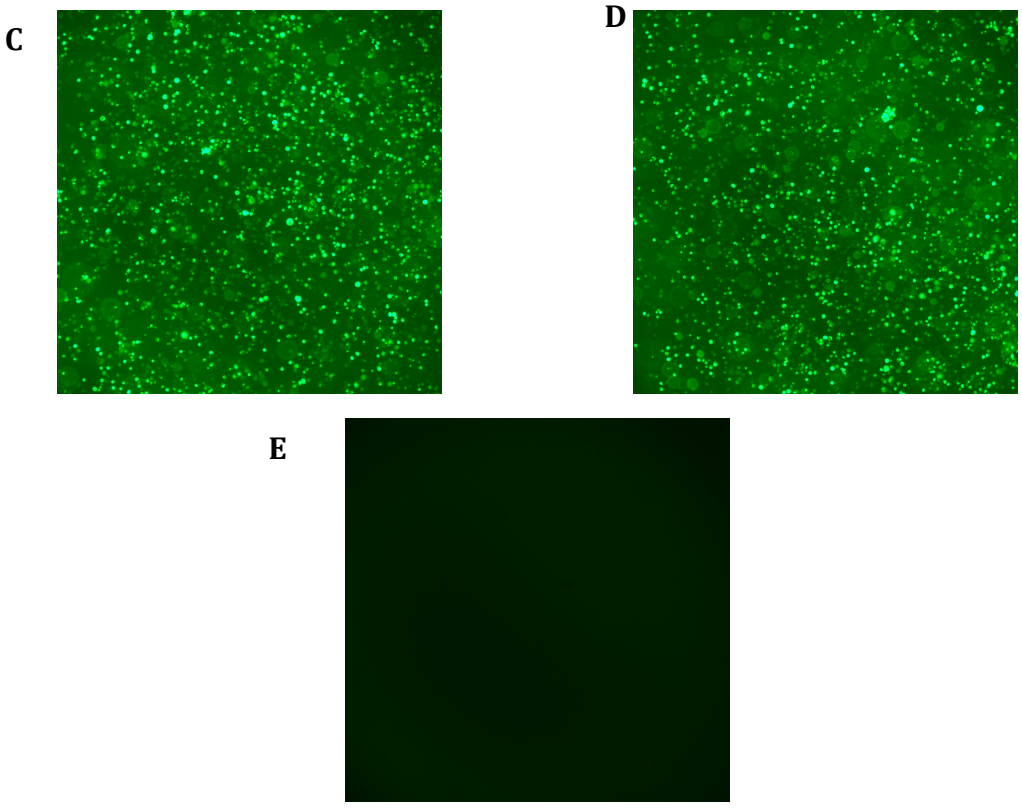

Figure 7 C, D, E: C) GFP Fluorescence at 96 hours post transfection D) Fluorescence at 120 hours post transfection. E) Untransfected cells

Western blots of cell lysates (**Figure 8**) at different time points show the expression of the gene over time. Western blots of media only showed hF9 secretion.

The purified protein was also treated with EndoH (an enzyme that removes high mannose-type N-glycans) and PNGaseF (an enzyme that removes all types of N-glycans including complex-type) to visualize the glycosylation patterns of smβ1-4GalNAcT to determine the whether or not the protein left the endoplasmic reticulum. Since EndoH and PNGaseF caused the same shift on a western blot, it appears that only high mannose-type N-glycans are found on smβ1-4GalNAcT (**Figure 8B**).

The cell lysates were also probed for LDN synthesis by western blot, but when the membrane was probed with an anti-LDN antibody, there does not appear to be any presence of LDN.

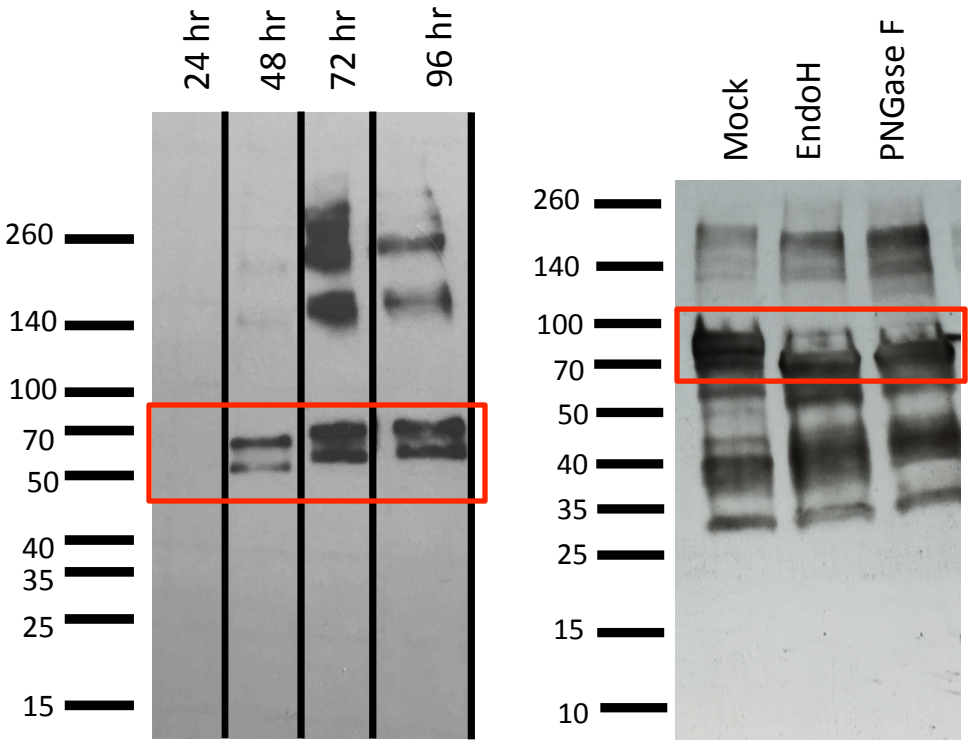

Figure 8: A) Non-reduced SDS-PAGE gel of hF9 (expression positive control) and smGalNAcT with an anti-GFP antibody. B) EndoH and PNGase treatment of smGalNAcT.

Activity Assays of Gene

The activity assays have shown that the gene is responsible for catalyzing the addition of a GalNAc to a terminal GlcNAc to form the LDN epitope. The acceptors used were a triose and chitobiose that were AEAB labeled and contained a terminal GlcNAc. The triose was derived from the tetrasaccharide LNnT and then treated with a beta-galactosidase to create a triose (trisaccharide) with a terminal GlcNAc. A time point assay was performed to see the activity of the gene over the course of a few days to ensure completion of reaction. The results of the reactions were measured using HPLC and MALDI-TOF Mass Spectrometry. The

results of the HPLC analysis are shown in **Figure 11** for the triose acceptor (GlcNAc-Gal-Glc-AEAB) and in **Figure 12** for the chitobiose acceptor (GlcNAc-GlcNAc-AEAB).

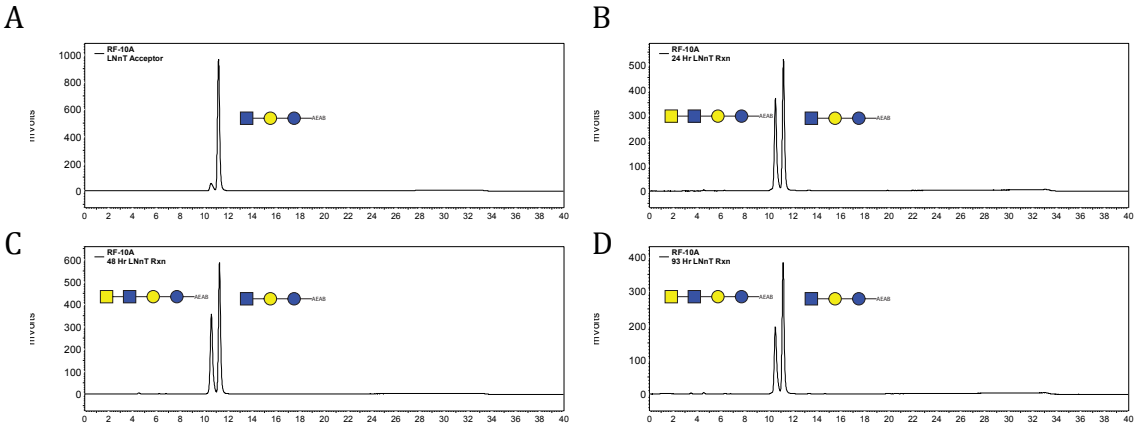

Figure 9: A) Triose acceptor B) Reaction after 24 hours C) Reaction after 48 hours D) Reaction after 93 hours

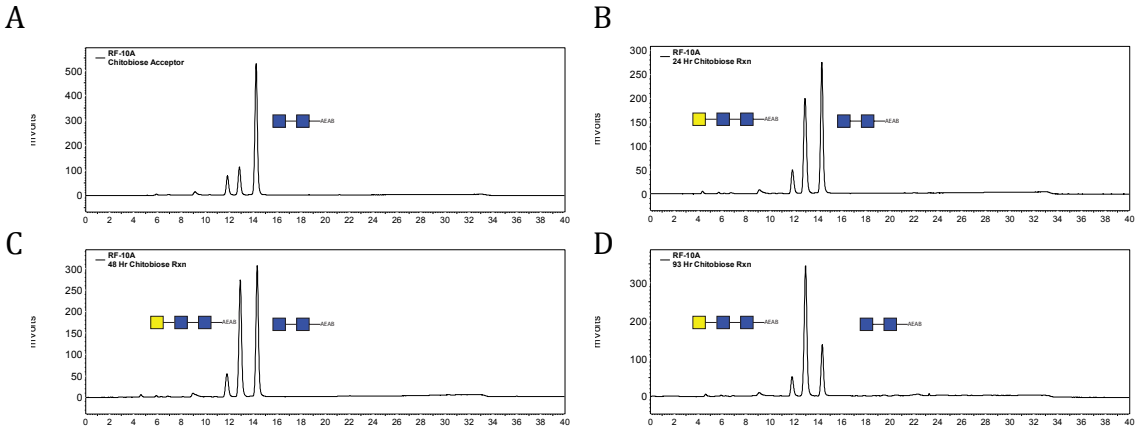

Figure 10: A) Chitobiose acceptor B) Reaction after 24 hours C) Reaction after 48 hours D) Reaction after 93 hours

A B

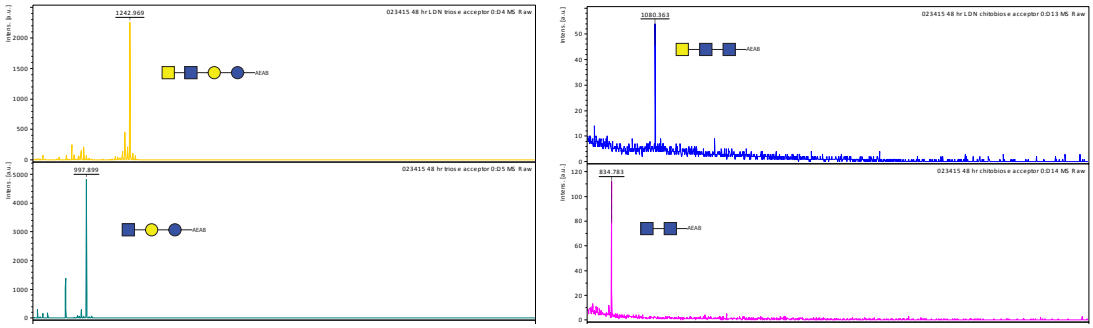

Figure 11: The two major peaks were collected from the 48 hour reaction with the triose (A) and chitobiose (B) acceptor. The molecular weights correspond to the peracetylated, permethylated carbohydrates with a sodium ion.

The data from the HPLC shows the generation of a new peak that runs off of the column earlier than the starting product. A time course that more clearly depicts the results is shown in **Figure 12**; the tallest peak in **12A** slowly diminish over time, replaced by a peak that elutes earlier and gets larger as the length of the reaction increases.

The peaks from the HPLC were collected, and then the glycans were peracetylated and permethylated. The peracetylated and permethylated carbohydrates from each peak were then analyzed by MALDI-TOF to obtain mass data. The representative mass data for the triose and chitobiose reactions are reported in **Figure 13 A** and **B**, respectively.

The results of other reactions found that the addition of EDTA inhibited transfer of UDP-GalNAc activity. Furthermore, smβ1-4GalNAcT cannot utilize UDP-Gal under the same conditions as UDP-GalNAc, at least with the two acceptors assayed.

Discussion

Based on the results provided, the gene Smp 021370 has been confirmed as a β1-4GalNAcTransferase. The purified smβ1-4GalNAcT is capable of synthesizing the LacDiNAc (GalNAcβ1-4GlcNAc-R, LDN) motif on at least two different acceptors, as well as retain activity for several days at 37°C. The ability to clone, express and characterize a gene from a human parasite is an important step in generating the necessary tools to understand parasite function at a molecular level. Furthermore, the data helps to show the strength of computational tools to predict gene products, as well as provide a new target for the development of pharmaceuticals. Furthermore, the characterization of the gene provides another link between the glycome and genome.

The protein alignments of the seven GalNAcTransferases annotated on GeneDB showed two distinct groups of sequences when using *C*. *elegans* β1-4GalNAcT as a reference sequence since it contains the necessary domains for glycosyltransferase functionality, as seen in **Figure 3A**. The four amino acid sequences that do not align well with ceGalNAcT (Smp 057620, Smp 159490, Smp 005500, Smp 139230) create a monophyletic group seen in **Figure 4**. Thus, it was deemed that those sequences were less likely to be β1-4GalNAcTs. There are other GalNAcTransferases, such as those that transfer a GalNAc directly to a peptide, which makes these four genes good candidates for exploring other GalNAcTransferase activities. Smp 021370 was a strong candidate for synthesis because not only did it align well; the sequence also was the most complete as compared to ceβ1-4GalNAcT. Furthermore, it shared an

isoleucine residue with ceβ1-4GalNAcT that was correlated with an increased preference in binding UDP-GalNAc (*51, 52*). A DVD sequence that corresponds to a metal binding site was also found in both ceβ1-4GalNAcT and smβ1-4GalNAcT when aligned. Since ceβ1-4GalNAcT used Mn^{2+}, it was predicted that smβ1-4GalNAcT would use the same metal ion cofactor. Lastly, Smp 021370 lacked a transmembrane domain, which was of little concern, since the goal was to have the smβ1-4GalNAcT secreted into the media for rapid purification. Furthermore, the close phylogenetic relationship between Smp 021370 and hB4GALNT3 provided extra support for β1-4GalNAcTransferase potential. Lastly, using the RPKM data, it appears that Smp 021370 is also the most expressed transcript among the three sequences that were the most closely related to ceβ1-4GalNAcT, making it the most biologically relevant transcript as well. Part of the goal is to create a drug target, thus, if this enzyme is the most highly expressed of the three potential β1-4GalNAcTransferases, then it would also represent the most rationale target since inhibition of this enzyme would also have the greatest effect.

Expression of the smβ1-4GalNAcT was interesting because, despite the lack of a predicted transmembrane domain, the protein appears to be found only in cell lysate and not secreted into the media. Post-translational modification of proteins by glycan maturation is often used to track protein movement through the secretory pathway. Sensitivity to EndoH only occurs for proteins that are found either in the endoplasmic reticulum or early in the Golgi complex. PNGaseF cleaves almost all known N-linked glycans (*55*). Since **Figure 8** shows the same shift if treated with

EndoH or PNGaseF it can be inferred that the smβ1-4GalNAcT was still early on in the secretory pathway, or the enzyme is resistant to typical processing of N-glycans leading to complex-type structures. This could be an interesting phenomenon that represents the interactions of the smβ1-4GalNAcT with another protein within the cell. However, it could also represent incorrect folding of the protein or an exposed sequence of the glycosyltransferase that prevents secretion. Furthermore, attempts to probe for *in vivo* activity of smβ1-4GalNAcT was unsuccessful using an anti-LDN antibody. The lack of LDN could be because the smβ1-4GalNAcT is not being exported out of the endoplasmic reticulum, and thus never interacts with an acceptor with a terminal GlcNAc. However, the lack of LDN motifs could also be due to the lack of terminal GlcNAcs because mammalian glycans are typically sialylated. Especially since ceβ1-4GalNAcT is secreted into the media, it leaves questions over what kind of intracellular interactions are occurring with smβ1-4GalNAcT. The Human Fut9, a glycosyltransferase used as an expression control, had no issues being secreted, which means that the signal sequence and GFP portion of the recombinant enzyme should not being causing issues with secretion.

The time courses shown in **Figures 9** and **10** provide important information, showing the conversion of the acceptor into the desired product with a terminal LDN epitope. Using a triose acceptor with a terminal GlcNAc (**Figure 9**), it appears that after 24 hours the entire product has been synthesized, and the reaction does not proceed with extended incubation. However, using a chitobiose acceptor, the smβ1-4GalNAcT proceeds to convert the acceptor into the final product

continuously over almost four days. The lifespan of the protein is remarkable, especially when it's being incubated at 37C. After the carbohydrates were peracetylated and permethylated, Mass data confirms the identity of the peaks shown in **Figures 9** and **10**. The molecular weights of peaks support the HPLC data since they primary peaks found in the purified fractions correspond to the molecular weight of the either the acceptor or product glycan.

Other activity assays were performed to elucidate other important reagents for the LDN synthesis reaction. Using EDTA, any metal co-factor was removed from the enzyme, which proceeded to inhibit activity of the glycosyltransferase. This information helps to support the need for $MnCl_2$ or another metal co-factor, which is supported computationally by the presence of a DVD (residues 180-182) and a DDD (residues 246-248) motif shared with ceβ1-4GalNAcT. These DxD motifs suggest binding to a metal cofactor. Due to similarities between the GalNAcTs and GalTs, UDP-Gal was also used in several activity assays in place of UDP-GalNAc. There was no synthesis of LN or any conversion of the acceptor, which leads to the theory that smβ1-4GalNAcT does not utilize UDP-Gal and is selective for the donor UDP-GalNAc. smβ1-4GalNAcT recognizes two different acceptors, which may show a proclivity for multiple acceptor sequences or demonstrate a basic requirement for a terminal GlcNAc with no preference for the preceding sequence. Lastly, the enzyme is functional a both 37°C and room temperature. The transcript is present in both the human and aquatic life stages (**Figure 5**), which supports the functionality of the enzymes at both temperatures.

Further experiments are necessary to glean more information about the enzyme. First and foremost, a survey of other acceptors is necessary, especially of biantennary glycans that is typical of many N-glycans. Activity assays with other ions, such as magnesium, could provide more detailed information on the ion-binding site. Other donors, such as UDP-GlcNAc and UDP-Glc, would also provide interesting information about the active site depending on the preference for the donors.

The ability to express the enzyme stably in a human cell line and be able to monitor the synthesis of LDN would assist in the development of the *S. mansoni* glycome on a human cell. Synthesis of LDN could be determined by flow cytometry by using an anti-LDN antibody previously characterized (*32*). While the ceβ1-4GalNAcT has been used previously to synthesize LDN on a human cell, it is possible that it is different from that created by a *S. mansoni* ortholog since *C. elegans* is a nematode and *S. mansoni* is a trematode.

Conclusion

LacDiNAc is an important motif in understanding the immune response to *S. mansoni*. A variety of antigens with LDN as a parent structure are presented by the parasite and work to modulate the immune system. The identification and characterization of enzymes important in the synthesis pathways of this important antigen could potentially lead to new therapeutics for schistosomiasis, which affects

hundreds of millions of people worldwide. The age of antibiotic resistance is rapidly approaching and has arrived for many diseases. In the arms race against resistance, it is necessary to not only develop new drugs, but also to discern new targets so that new classes of pharmaceuticals can be developed.

A novel β1-4GalNAcTransferase from an important agent of a disease with a global impact has been cloned, expressed, and characterized. The work here represents the first β1-4GalNAcTransferase expressed from a trematode. Schistosomiasis is a neglected tropical disease with little research focused on the disease. Partially it is because there is little financial gain to be made from studying this parasite. However, there are global consequences for the advances in understanding the worm biochemistry because it could potentially lead to vaccines and novel treatments. Thus, it is important that strides are made in the research of the molecular biology of parasites because of the little attention neglected tropical diseases receive.

With the advent of next generation sequencing technologies, the ability to sequence the entire genome of any organism becomes a very real scenario. The genome of an organism is an incredible tool to have in the identification of genes and proteins important in the life cycle of an organism. Especially in the realm of multicellular parasites, the ability to synthesize gene products *de novo* simplifies the process of identifying and cloning these important gene products. The computational tools available now are sophisticated and are able to correctly identify certain conserved

motifs in DNA and amino acids sequences, but the information is still limited. In the understudied field of glycobiology, there is little information on the different glycosyltransferases outside of select model organisms, especially when it comes to understanding binding affinities for acceptors. While human glycosyltransferases have been studied much more extensively, this work represents the first GalNAcTransferase expressed from any of the *Schistosoma* species. A BlastP search of the amino acid sequence of Smp 021370 shows a sequence of homology that is identified as being an UDP-Gal binding site. However, as shown in the data provided, the smβ1-4GalNAcT does not utilize UDP-Gal under the same conditions that it utilizes UDP-GalNAc. Therefore, it is important to confirm computational annotations of genes with physical data. Furthermore, physical data provides much more insight into an amino acid sequence by elucidating the donor and acceptor combinations to understand the specificity of the enzyme in question. Other questions, such as temperature dependence and kinetics have yet to be predicted computationally. However, the data presented in this study could help to not only help prediction models for glycosyltransferases, but also helps to link the glycome and the genome. Predicting glycosylation patterns is notoriously difficult, and any information on the glycosyltransferases necessary for glycan synthesis would help in developing glycosylation prediction models, such as Structure Feature Analysis Tool (*56*) or Glycosylation Predictor (*57*). Since many proteins and lipids are glycosylated, it is necessary to understand glycan modification of molecules to understand the interactions occurring *in vivo*. Furthermore, having the actual protein allows for other experiments to screen for inhibitors of the enzyme and find

potential new therapeutics. In the case of *S. mansoni*, this is particularly important since there still remains only one pharmaceutical capable of clearing the helminth after infection. Hopefully, the data presented here represents that first step towards identifying new drug targets for schistosomiasis.

References

1. W. H. Organization. (2014).
2. A. G. Ross *et al.*, Schistosomiasis. *The New England journal of medicine* **346**, 1212 (Apr 18, 2002).
3. C. H. King, in *The Causes and Impacts of Neglected Tropical and Zoonotic Diseases: Opportunities for Integrated Intervention Strategies*. (Washington (DC), 2011).
4. A. Fenwick *et al.*, The Schistosomiasis Control Initiative (SCI): rationale, development and implementation from 2002-2008. *Parasitology* **136**, 1719 (Nov, 2009).
5. A. Fenwick, J. P. Webster, Schistosomiasis: challenges for control, treatment and drug resistance. *Current opinion in infectious diseases* **19**, 577 (Dec, 2006).
6. M. J. Doenhoff, D. Cioli, J. Utzinger, Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis. *Current opinion in infectious diseases* **21**, 659 (Dec, 2008).
7. M. H. Bonds, D. C. Keenan, P. Rohani, J. D. Sachs, *Poverty trap formed by the ecology of infectious diseases*. (2010), vol. 277, pp. 1185-1192.
8. Prevention and control of schistosomiasis and soil-transmitted helminthiasis. *World Health Organization technical report series* **912**, i (2002).
9. V. S. Files, E. B. Cram, A Study on the Comparative Susceptibility of Snail Vectors to Strains of Schistosoma mansoni. *The Journal of Parasitology* **35**, 555 (1949).
10. J. R. Ingram *et al.*, Investigation of the Proteolytic Functions of an Expanded Cercarial Elastase Gene Family in <italic>Schistosoma mansoni</italic>. *PLoS Negl Trop Dis* **6**, e1589 (2012).
11. J. McKerrow, J. Salter, Invasion of skin by Schistosoma cercariae. *Trends in Parasitology* **18**, 193 (5/1/, 2002).
12. M. A. Stirewalt, in *Advances in Parasitology,* D. Ben, Ed. (Academic Press, 1974), vol. Volume 12, pp. 115-182.
13. W. Haas, R. Schmitt, Characterization of chemical stimuli for the penetration of Schistosoma mansoni cercariae. I. Effective substances, host specificity. *Zeitschrift fur Parasitenkunde (Berlin, Germany)* **66**, 293 (1982).
14. M. K. Jones, S. Lustigman, A. Loukas, Tracking the Odysseys of Juvenile Schistosomes to Understand Host Interactions. *PLoS Neglected Tropical Diseases* **2**, e257 (07/16, 2008).
15. A. W. Cheever, J. G. Macedonia, J. E. Mosimann, E. A. Cheever, Kinetics of egg production and egg excretion by Schistosoma mansoni and S. japonicum in mice infected with a single pair of worms. *The American journal of tropical medicine and hygiene* **50**, 281 (Mar, 1994).
16. E. J. Pearce, A. S. MacDonald, The immunobiology of schistosomiasis. *Nat Rev Immunol* **2**, 499 (07//print, 2002).
17. T. Elbaz, G. Esmat, Hepatic and Intestinal Schistosomiasis: Review. *Journal of Advanced Research* **4**, 445 (9//, 2013).
18. C. f. D. Control. (2013).

19. B. Gryseels, K. Polman, J. Clerinx, L. Kestens, Human schistosomiasis. *Lancet* **368**, 1106 (Sep 23, 2006).
20. N. M. Nour, Schistosomiasis: Health Effects on Women. *Reviews in Obstetrics and Gynecology* **3**, 28 (Winter, 2010).
21. J. F. Doherty, A. H. Moody, S. G. Wright, *Lesson of the Week: Katayama fever: an acute manifestation of schistosomiasis*. (1996), vol. 313, pp. 1071-1072.
22. C. H. King, M. Dangerfield-Cha, The unacknowledged impact of chronic schistosomiasis. *Chronic Illness* **4**, 65 (March 1, 2008, 2008).
23. A. M. Agnew, H. M. Murare, M. J. Doenhoff, Immune attrition of adult schistosomes. *Parasite immunology* **15**, 261 (May, 1993).
24. A. R. de Jesus *et al.*, Clinical and Immunologic Evaluation of 31 Patients with Acute Schistosomiasis mansoni. *Journal of Infectious Diseases* **185**, 98 (January 1, 2002, 2002).
25. J. A. Jackson, I. M. Friberg, S. Little, J. E. Bradley, Review series on helminths, immune modulation and the hygiene hypothesis: Immunity against helminths and immunological phenomena in modern human populations: coevolutionary legacies? *Immunology* **126**, 18 (10/08/received
10/08/revised
10/08/accepted, 2009).
26. R. M. Maizels, M. Yazdanbakhsh, Immune regulation by helminth parasites: cellular and molecular mechanisms. *Nat Rev Immunol* **3**, 733 (Sep, 2003).
27. R. T. DAMIAN, Parasite immune evasion and exploitation: reflections and projections. *Parasitology* **115**, 169 (1997).
28. R. T. Damian, Molecular Mimicry in Biological Adaptation. *Science* **147**, 824 (Feb 19, 1965).
29. I. van Die, R. D. Cummings, Glycan gimmickry by parasitic helminths: A strategy for modulating the host immune response? *Glycobiology* **20**, 2 (January 1, 2010, 2010).
30. S. Meyer, B. Tefsen, A. Imberty, R. Geyer, I. van Die, The C-type lectin L-SIGN differentially recognizes glycan antigens on egg glycosphingolipids and soluble egg glycoproteins from Schistosoma mansoni. *Glycobiology* **17**, 1104 (October 1, 2007, 2007).
31. C. K. Makaaru, R. T. Damian, D. F. Smith, R. D. Cummings, The human blood fluke Schistosoma mansoni synthesizes a novel type of glycosphingolipid. *Journal of Biological Chemistry* **267**, 2251 (February 5, 1992, 1992).
32. A. K. Nyame, A. M. Leppanen, B. J. Bogitsh, R. D. Cummings, Antibody responses to the fucosylated LacdiNAc glycan antigen in Schistosoma mansoni-infected mice and expression of the glycan among schistosomes. *Experimental parasitology* **96**, 202 (Dec, 2000).
33. K.-H. Khoo *et al.*, A Unique Multifucosylated −3GalNAc14GlcNAc13Gal1- Motif Constitutes the Repeating Unit of the Complex O-Glycans Derived from the Cercarial Glycocalyx of Schistosoma mansoni. *Journal of Biological Chemistry* **270**, 17114 (July 21, 1995, 1995).
34. N. S. Prasanphanich, M. L. Mickum, J. Heimburg-Molinaro, R. D. Cummings, Glycoconjugates in host-helminth interactions. *Frontiers in Immunology* **4**, (2013-August-28, 2013).

35. N. S. Prasanphanich *et al.*, Immunization with recombinantly expressed glycan antigens from Schistosoma mansoni induces glycan-specific antibodies against the parasite. *Glycobiology*, (April 11, 2014, 2014).
36. K. K. Van de Vijver, A. M. Deelder, W. Jacobs, E. A. Van Marck, C. H. Hokke, LacdiNAc- and LacNAc-containing glycans induce granulomas in an in vivo model for schistosome egg-induced hepatic granuloma formation. *Glycobiology* **16**, 237 (March 1, 2006, 2006).
37. A. van Remoortere *et al.*, Profiles of Immunoglobulin M (IgM) and IgG Antibodies against Defined Carbohydrate Epitopes in Sera of Schistosoma-Infected Individuals Determined by Surface Plasmon Resonance. *Infection and Immunity* **69**, 2396 (April 1, 2001, 2001).
38. E. van Liempt *et al.*, Specificity of DC-SIGN for mannose- and fucose-containing glycans. *FEBS letters* **580**, 6123 (Nov 13, 2006).
39. M. H. Meevissen *et al.*, Specific glycan elements determine differential binding of individual egg glycoproteins of the human parasite Schistosoma mansoni by host C-type lectin receptors. *International journal for parasitology* **42**, 269 (2012).
40. A. van Diepen *et al.*, Differential anti-glycan antibody responses in Schistosoma mansoni-infected children and adults studied by shotgun glycan microarray. *PLoS Negl Trop Dis* **6**, e1922 (2012).
41. A. Zerlotini *et al.*, SchistoDB: a Schistosoma mansoni genome resource. *Nucleic acids research* **37**, D579 (January 1, 2009, 2009).
42. A. V. Protasio *et al.*, A systematically improved high quality genome and transcriptome of the human blood fluke Schistosoma mansoni. *PLoS Negl Trop Dis* **6**, e1455 (Jan, 2012).
43. M. L. Mickum, N. S. Prasanphanich, J. Heimburg-Molinaro, K. E. Leon, R. D. Cummings, Deciphering the glycogenome of schistosomes. *Frontiers in Genetics* **5**, 262 (08/05
05/31/received
07/15/accepted, 2014).
44. J. Rini, J. Esko, A. Varki, in *Essentials of Glycobiology,* A. Varki *et al.*, Eds. (Cold Spring Harbor (NY), 2009).
45. Z. S. Kawar, I. Van Die, R. D. Cummings, Molecular Cloning and Enzymatic Characterization of a UDP-GalNAc:GlcNAcβ-R β1,4-N-Acetylgalactosaminyltransferase from Caenorhabditis elegans. *Journal of Biological Chemistry* **277**, 34924 (September 20, 2002, 2002).
46. H. Blum, H. Beier, H. J. Gross, Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *ELECTROPHORESIS* **8**, 93 (1987).
47. X. Song *et al.*, Novel fluorescent glycan microarray strategy reveals ligands for galectins. *Chem Biol* **16**, 36 (Jan 30, 2009).
48. X. Song, J. Heimburg-Molinaro, D. F. Smith, R. D. Cummings, Derivatization of free natural glycans for incorporation onto glycan arrays: derivatizing glycans on the microscale for microarray and other applications (ms# CP-10-0194). *Current protocols in chemical biology* **3**, 53 (2011).

49. A. J. Hanneman, J. C. Rosa, D. Ashline, V. N. Reinhold, Isomer and glycomer complexities of core GlcNAcs in Caenorhabditis elegans. *Glycobiology* **16**, 874 (September 1, 2006, 2006).
50. P. Kang, Y. Mechref, I. Klouckova, M. V. Novotny, Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid communications in mass spectrometry : RCM* **19**, 3421 (2005).
51. B. Ramakrishnan, P. K. Qasba, Structure-based Design of β1,4-Galactosyltransferase I (β4Gal-T1) with Equally EfficientN-Acetylgalactosaminyltransferase Activity: POINT MUTATION BROADENS β4Gal-T1 DONOR SPECIFICITY. *Journal of Biological Chemistry* **277**, 20833 (June 7, 2002, 2002).
52. T. Sato *et al.*, Molecular Cloning and Characterization of a Novel Human β 1,4-N-Acetylgalactosaminyltransferase, β 4GalNAc-T3, Responsible for the Synthesis of N,N′ -Diacetyllactosediamine, GalNAc β 1‑4GlcNAc. *Journal of Biological Chemistry* **278**, 47534 (November 28, 2003, 2003).
53. Y. Ikehara *et al.*, Apical Golgi localization of N,N'-diacetyllactosediamine synthase, beta4GalNAc-T3, is responsible for LacdiNAc expression on gastric mucosa. *Glycobiology* **16**, 777 (Sep, 2006).
54. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **5**, 621 (Jul, 2008).
55. H. H. Freeze, C. Kranz, Endoglycosidase and Glycoamidase Release of N-Linked Glycans. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **0 17**, 10.1002/0471142727.mb1713as89 (2010).
56. P. V. Lam *et al.*, Structure-based comparative analysis and prediction of N-linked glycosylation sites in evolutionarily distant eukaryotes. *Genomics, proteomics & bioinformatics* **11**, 96 (Apr, 2013).
57. S. E. Hamby, J. D. Hirst, Prediction of glycosylation sites using random forests. *BMC bioinformatics* **9**, 500 (2008).