Christopher Taus, MSc

Identification, molecular cloning and characterization of the first mollusc glycosyltransferase involved in O-glycan biosynthesis

Dissertation

zur Erlangung des akademischen Grades
„Doktor nat. techn.“
an der Universität für Bodenkultur, Wien

Betreuung:
Univ.-Prof. Dr. Erika Staudacher (Dep. für Chemie)
Univ.-Prof. Dr. Reingard Grabherr (Dep. für Biotechnologie)

2013

ABSTRACT

The aim of this work was to isolate and characterize novel glycosyltransferases from snail origin, especially from *Biomphalaria glabrata*, which is the intermediate host of the human parasite *Schistosoma mansoni*.

Gastropods have an extraordinary glycosylation potential with some uncommon carbohydrate modifications, such as methylation and ß1,2-xylosylation. The latter glycan feature occurs typically in plants, but is very uncommon in the animal kingdom. For this reason, the ß1,2-xylosyltransferase (ß1,2-XylT; EC 2.4.2.38) was one desired enzyme to characterize the first time from gastropods. However, creation of expression cDNA libraries in insect cells and subsequent cell sorting procedures, based on ß1,2-xylosylated N-glycans exposed on the cell surface, did not result in the isolation of a snail ß1,2-XylT gene. As a positive control, it was shown that N-glycans of membrane proteins in insect cells were ß1,2-xylosylated, as soon as a recombinant ß1,2-XylT was expressed from *Arabidopsis thaliana*. The cell sorting strategy was based on these data. In addition, homology-based PCR and biochemical purification failed the isolation of the first ß1,2-XylT from gastropods.

Based on previous studies, gastropods synthesize mucin-type O-glycans, which play important roles in host finding processes, such as between *S. mansoni* and *B. glabrata*. In this study the initiating enzyme, a polypeptide *N*-acetylgalactosaminyltransferase (ppGalNAcT; EC 2.4.1.41), was identified and characterized. By homology-based PCR and subsequent RACE PCR(s) the full-length gene of this enzyme from *B. glabrata* could be isolated from a cDNA library, which was created in this work. This study represents the first recombinant expression and fully characterization of this enzyme family from mollusk kingdom. According to its structure and its substrate specificity it is a close relative to ppGalNAcT-2 isoforms from other animal species.

Furthermore, the core-1 ß1,3-galactosyltransferase (core-1 ß1,3-GalT or T-Synthase; EC 2.4.1.122) responsible for the synthesis of the core-1 O-glycan (T-antigen), was isolated to almost 50% from the *B. glabrata* cDNA library. Using homology-based PCR and a subsequent RACE PCR, the complete 5'end was identified and showed characteristics of a type II membrane protein, as well as had high homology to T-synthases from other organisms.

By a screening procedure of the cDNA library from *Arion lusitanicus*, a full-length C-type lectin was identified and so far expressed recombinantly in *Escherichia coli*, *Pichia pastoris* and insect cells.

ZUSAMMENFASSUNG

Das Ziel dieser Forschungsarbeit war die Isolierung und die folgende biochemische Charakterisierung neuer Glykosyltransferasen von Schnecken, speziell von *Biomphalaria glabrata*, die den Zwischenwirt von *Schistosoma mansoni* darstellt.
Schnecken haben ein außergewöhnliches Glykosylierungsmuster, wobei Methylierungen und die ß1,2-Xylosylierung von N-Glykanen eine der interessantesten Modifikationen darstellen. Letztere Eigenschaft ist typisch für pflanzliche N-Glykane, jedoch im tierischen Reich sehr ungewöhnlich. Nicht zuletzt aus diesem Grund lag ein Augenmerk dieser Dissertation auf der Charakterisierung einer ß1,2-Xylosyltransferase (ß1,2-XylT; EC 2.4.2.38). Die Erstellung von cDNA Bibliotheken, die mittels FACS gesortet wurden, konnte keine Zellen, die die ß1,2-XylT Gensequenz beinhalteten, isolieren. Diese Strategie der Genisolierung begründete sich auf der Charakterisierung der Positivkontrolle, in der gezeigt werden konnte, dass N-Glykane von Membranproteinen ß1,2-xylosyliert werden, sobald eine rekombinante ß1,2-XylT (*Arabidopsis thaliana*) exprimiert wurde. Homologie-PCRs und Proteinreinigungen führten ebenso zu keiner erfolgreichen Isolierung der ersten ß1,2-Xylosyltransferase aus Schnecken.
Basierend auf Daten früherer Studien synthetisieren Gastropoden Mucin-Typ O-Glykane, die Wirts-Parasit-Interaktionen beeinflussen können. Im Laufe dieser Studie konnte eine Polypeptid *N*-Acetylgalaktosaminyltransferase (ppGalNAcT; EC 2.4.1.41) identifiziert und charakterisiert werden. Durch Homologie-PCR und darauf aufbauenden RACE-PCRs konnte die vollständige Gensequenz der ersten ppGalNAcT von *B. glabrata* aus einer, in dieser Arbeit erstellten, cDNA Bibliothek isoliert werden. Die dadurch ermöglichte, folgende Charakterisierung stellt die erste Beschreibung dieser Enzymfamilie aus Mollusken dar. Bezogen auf die Struktur und die Substratsspezifität ist die hier isolierte ppGalNAcT nahe zu T-2 Isoformen anderer Spezies verwandt.
Weiters konnte zu 50% die Gensequenz einer Core-1 ß1,3-Galaktosyltransferase (Core-1 ß1,3-GalT oder T-Synthase; EC 2.4.1.122) von *B. glabrata* identifiziert werden, die für die Synthese des Core-1 O-Glykans (T-Antigen) verantwortlich ist. Durch Homologie- und RACE-PCR konnte das vollständige 5’ Ende des Gens

analysiert werden, das Eigenschaften gewöhnlicher Typ II Mempranproteine aufweist und weiters hohe Homologie zu anderen T-Synthasen hat.

Im Rahmen einer detailierten Qualitätskontrolle der cDNA Bibliothek von *A. lusitanicus*, konnte das vollständige Gen eines C-Typ Lektins identifiziert werden. Bisher wurde das Protein in *Escherichia coli*, *Pichia pastoris* and Insektenzellen rekombinant hergestellt.

DANKSAGUNG

An dieser Stelle möchte ich einigen Personen aus tiefstem Herzen Danke sagen, deren Unterstützung, Ratschläge und Diskussionen zu der Vollendung dieser Arbeit beigetragen haben. Bedanken will ich mich bei…..

…meiner Betreuerin, Univ.-Prof. Dr. Erika Staudacher, die es mir ermöglichte in Ihrem Labor, als auch in Japan, zu forschen, mich mit diesem spannenden Thema der Glykobiologie auseinanderzusetzen, mir enormes Vertrauen bezüglich der Arbeitsplanung entgegenbrachte bzw. stets ein offenes Ohr für etwaige Probleme hatte.

…meiner zweiten Betreuerin, Univ.-Prof. Dr. Reingard Grabherr, die mich im Bereich der Molekularbiologie professionell unterstützte und sich immer Zeit für hilfreiche Diskussionen nahm.

…Prof. Kiyoshi Furukawa und Dr. Takeshi Sato für die Möglichkeit in Japan (Nagaoka) zu arbeiten und natürlich für all die gemeinsamen Aktivitäten und wunderschöne Zeit.

…Doz. DI Dr. Richard Strasser, Univ.-Prof. Dr. Lukas Mach, Dr. Michael Sauer und Dr. Hans Marx für die unproblematische und freundliche Bereitstellung verschiedenster Proben bzw. Laborgeräte.

…Dr. Martin Dragosits für Modellierungsstudien und Dipl. Ing. Markus Windwarder für MS-Messungen.

…Mag. Laura Neumann, Dr. Dieter Palmberger und Stefan Gross für hilfreiche Diskussionen im Laufe meiner Doktorarbeit.

…all meinen restlichen Arbeitskollegen am Department für Chemie, sowie am Department für Biotechnologie, für diese unvergessliche, lustige Zeit, die ich sehr vermissen werde.

…nicht zuletzt meiner Familie und meinen Freunden, deren Unterstützung sehr hilfreich für die Beendigung dieser Doktorarbeit war.

CONTENTS

1 INTRODUCTION

1.1 Importance of glycosylation

Proteins undergo different types of covalent modification with glycosylation as the most common one in living organisms (Lis & Sharon, 1993). Also archaea-, as well as eubacteria produce glycoproteins, even though structures are different to higher organisms (Gerwig et al, 1992; Lechner & Wieland, 1989; Messner & Sleytr, 1991). Glycans differ not only in sequence and chain length, but also in anomery, position of linkages, branching points of monosaccharide attachments and further modifications, such as sulphate, phosphate, acetyl and methyl groups. Glycosylation is organism, cell and tissue specific resulting in microheterogeneity of glycoproteins. Carbohydrates influence protein features, such as correct folding and protection against proteolysis (Lis & Sharon, 1993), as well as have several biological functions. In terms of the latter, carbohydrate chains act as recognition determinants in a variety of physiological and pathological processes (Brandley & Schnaar, 1986; Rademacher et al, 1988; Sharon & Lis, 1993), which include clearance of glycoproteins (Ashwell & Harford, 1982), protein trafficking (Kornfeld, 1992) and a wide range of cell-cell interactions. The last topic ranges from the attachment process of sperm and ova (Wassarman, 1990) to adhesion of microorganisms to host tissue (Leffler & Svanborg-Eden, 1990; Ofek & Sharon, 1988).

1.2 N- and O-glycosylation

Glycoconjugates can be linked to proteins by an amide bond of Asn (N-glycans), or linked via oxygen of Ser, Thr, Tyr or hydroxyproline (Hyp) (O-glycans). Besides the linkage of glycans to a polypeptide chain, these molecules are linked to lipids, so called glycolipids, which are classified into two subgroups, glycosphingolipids and phosphoglycolipids.

N-glycosylation occurs in all eukaryotic cells. The biosynthesis can be divided into two main steps, the assembly of the lipid-linked glycan located in the endoplasmic membrane and the transfer of this oligosaccharide to selected Asn residues of a nascent polypeptide chain. Synthesis of dolichol-linked oligosaccharide is dependent on the availability of three different nucleotide activated sugar donors, namely UDP-*N*-acetylglucosamine (UDP-GlcNAc), GDP-mannose (GDP-Man) and UDP-glucose (UDP-Glc). The biosynthesis of N-glycans is initiated at the cytoplasmic side of the ER membrane by addition of a GlcNAc-phosphate residue to dolichol phosphate. The enzyme catalyzing this reaction, *N*-Acetylglucosaminylphosphate-transferase, can be inhibited by tunicamycin. The next step of N-glycan biosynthesis is the addition of a second GlcNAc from the activated sugar UDP-GlcNAc. Several mannosyltransferases, some of them characterized in detail from *Saccharomyces cerevisiae* (Huffaker & Robbins, 1983) assemble the glycan $Man_5GlcNAc_2$-PP-Dol. Following a flip of this precursor glycan to the luminal side of the ER, four mannosyltransferases and three glucosyltransferases are involved in the completion of the lipid-linked core oligosaccharide $Glc_3Man_9GlcNAc_2$. The synthesis in the ER lumen occurs in a stepwise manner (Burda et al, 1996; Rearick et al, 1981). The use of dolichol pyrophosphate (Dol-PP) as a carrier for glycan assembly and the transfer of $Glc_3Man_9GlcNAc_2$ to specific Asn within the consensus sequence Asn-X-Ser/Thr (X can be every amino acid except Pro (Gavel & von Heijne, 1990)) is highly conserved in eukaryotes (Burda & Aebi, 1999).

Mature N-glycans are classified into three subgroups, oligo- (high-) mannosidic-, hybrid- and complex type ones (Figure 1). All of these N-glycan classes have a common core structure, consisting of two inner GlcNAc residues and three Man residues.

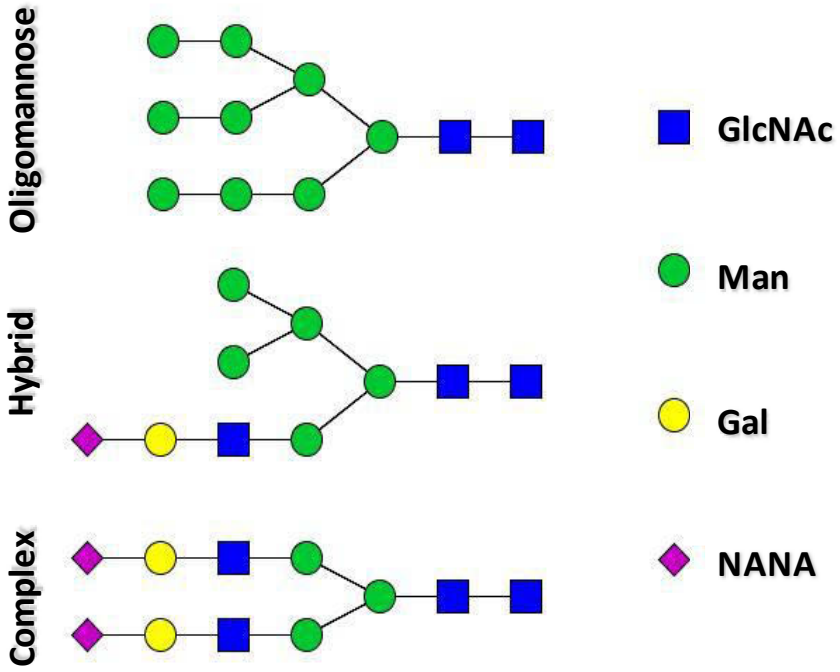

Figure 1: Types of N-glycans

Following the covalent attachment of the sugar to the protein, processing of the novel glycan first occurs in the ER by sequential removal of glucose residues by α-glucosidases I and II. Most glycoproteins are modified by ER α-mannosidase I yielding in a $Man_8GlcNAc_2$ isomer. In the Golgi of multicellular organisms, the removal of Man continues by the activity of α1,2-mannosidases resulting in $Man_5GlcNAc_2$, which is a key intermediate in the pathway to hybrid and complex N-glycans. However, not all N-glycans are processed in this way, resulting in oligo- or high-mannosidic (e.g. yeast) structures.

In the medial Golgi *N*-acetylglucosaminyltransferase I (GlcNAcT-I) and α-mannosidase II are synthesizing $GlcNAcMan_3GlcNAc_2$. GlcNAcT-II activity, which adds a GlcNAc to the α1,6-linked Man, provides the precursor for all biantennary, complex N-glycans. Hybrid N-glycans can be produced, if the α-mannosidase II is not acting on the glycan. Truncated N-glycans, so called paucimannosidic structures, are found in high amounts in invertebrates and plants. In this case, a Golgi hexosaminidase that acts after α-mannosidase II is responsible for producing $Man_{3-4}GlcNAc_2$.

Glycosyltransferases, localized in the trans Golgi, add different sugars to the processed N-glycan resulting in a great variety of mature, complex N-glycans. Three major ways are possible, namely (1) core modification, (2) glycan elongation and (3) "capping" of elongated branches. Later on, the focus is on modification of the N-glycan core by adding different sugars. Whereas in vertebrates, fucosylation occurs only α1,6-linked to the GlcNAc adjacent to the Asn, invertebrates additionally fucosylate in an α1,3-linkage. Invertebrates, such as insects, exhibit glycans in a di-fucosylated state too. Plants only have α1,3-fucosylated N-glycans. Besides fucosylation, another common core modification of plants and some invertebrates, such as molluscs, is ß1,2-xylosylation of the ß-Man. Both, core fucosyltransferases and the ß1,2-XylT require in most cases the prior action of GlcNAcT-I for their activity (Brockhausen et al, 2008).
Because of the missing ß1,2-Xyl and α1,3-Fuc in mammalian N-glycans, these epitopes are known to be important IgE binding carbohydrate determinants of plant allergens (Aalberse et al, 1981; van Ree et al, 2000; Wilson & Altmann, 1998).

The O-glycosylation linkage of proteins is much more diverse than the N-glycosylation linkage. In humans seven different types of O-glycans have been identified, such as (1) mucin-type, (2) glycosaminoglycans, (3) O-linked GlcNAc, (4) O-linked galactose (Gal), (5) O-linked Man (6) O-linked Glc (7) O-linked fucose (Fuc) (Wopereis et al, 2006). Furthermore, every amino acid with a hydroxyl functional group, such as Ser, Thr, Tyr, Hyp can be used as glycosylation site (Spiro, 2002). The most common O-glycosylation type is formed by the addition of *N*-acetylgalactosamine (GalNAc) to Ser/Thr residues, the so called mucin-type O-glycosylation. The other types are rare or restricted to specific species, tissues or proteins. For example, O-linked GlcNAc has been detected in cytosolic and nuclear proteins (Hart et al, 1989), Fucα-*O*-Ser was found in epidermal growth factor domains (Harris & Spellman, 1993), Man-*O*-Ser was identified in yeast cells (Duman et al, 1998; Mormeneo et al, 1989) and O-linked Gal was detected in collagens. Although, there are algorithms to predict O-glycosylation sites, no defined consensus sequence has been identified so far (Hansen et al, 1995). Here, only mucin-type O-glycosylation will be discussed in detail, especially the enzymes responsible for the synthesis of the so called T-antigen.

Besides mammals, mucin-type O-glycosylation has been found in insects, snails, worms, and various parasites (Wilson et al, 2009; Stepan et al, 2012). Also higher plants are able to initiate glycosylation of their proteins by GalNAc transfer to Ser or Thr (Kishimoto et al, 1999). This protein modification is restricted to proteins that pass the Golgi compartment. At least eight different core structures of mucin-type O-glycans occur in mammalian glycoproteins (Figure 2).

The initiating glycosyltransferases for this type of glycosylation are members of an evolutionarily conserved family, the so called polypeptide *N*-acetylgalactosaminyltransferases (ppGalNAcTs, [EC 2.4.1.41]). These enzymes transfer a GalNAc residue from UDP-GalNAc to a Ser or Thr within a polypeptide chain. Depending on the organism, tissue and developmental stage, the protein bound GalNAc residue is modified by other monosaccharides. First a GlcNAc and/or Gal residue is linked to the GalNAc and subsequently it can be modified by more of these or other sugars. Mucin-type O-glycosylation has been suggested to influence conserved processes in development. In addition, alterations in O-glycosylation patterns have been connected with pathogenic events, especially with cancer progression and metastasis.

Up to now, 20 representatives of ppGalNAcTs have been identified in humans and several more in other organisms. According to conserved domains, intron/exon position and their acceptor preferences, they have been characterized and grouped into subfamilies (Bennett et al, 2012). Some common structural elements have been described for these enzymes. They have a membrane spanning domain at the N-terminus, which is typical for type II membrane proteins and are located in the Golgi apparatus. Furthermore, ppGalNAcTs share a highly conserved UDP-GalNAc binding region, a manganese-binding site and a lectin-like domain located at the C-terminus, which is important for acceptor substrate specificity (Raman et al, 2012). In addition to mammalian ones, a few invertebrate ppGalNAcTs have been characterized in terms of their specificity and biochemical properties. Within the *Drosophila melanogaster* genome 14 members of this enzyme family are encoded with some of them being important for normal development (Ten Hagen et al, 1998; Ten Hagen & Tran, 2002). *Caenorhabditis elegans* contains 11 isoforms of ppGalNAcTs (Hagen & Nehrke, 1998). So far, some parasites were also described to be able for synthesizing mucin-type O-glycosylation, such as *Fasciola hepatica* (Freire et al, 2003a)*, Trypanosoma cruzi* (Freire et al, 2003b) *and Toxoplasma gondii*

(Stwora-Wojczyk et al, 2004). The review published from Bennett (Bennett et al, 2012) gives a comprehensive coverage of the ppGalNAcTs that have been discovered so far in different organisms.

Core 1 ß-1,3-galactosyltransferase, also known as T-synthase, is the key enzyme for the synthesis of the common core 1 O-glycan structure (T-antigen). This enzyme family transfers Gal from UDP-Gal to the GalNAc residue of the Tn-antigen (GalNAcα1-Ser/Thr), resulting in the T-antigen (Galß1-3GalNAcα1-Ser/Thr). This glycan structure is the most common precursor for a variety of complex structures of mucin-type O-glycans, normally found in mucins, membrane glycoproteins and secreted glycoproteins. Whereas in human there is only one active T-synthase identified so far (Ju et al, 2002), in the model organism *D. melanogaster* 4 genes are coding for active T-synthases (Müller et al, 2005). In *C. elegans* so far one gene, which showed enzyme activity, was found that is highly homologue to human T-synthase (Ju et al, 2006).

Whereas invertebrate T-synthases are N-glycosylated, mammalian ones do not have this modification, so in their case the chaperone Cosmc is required for its functional activity (Ju & Cummings, 2002). Core 2 (Figure 2) is derived from core 1 by the additional linkage of a GlcNAc residue in ß1,6-linkage to the protein bound GalNAc. Core 3 and core 4 (Figure 2) exhibit a more restricted, tissue specific expression pattern and occurs in bronchial mucins (Breg et al, 1988), colonic secretions (Podolsky, 1985) and fetal mucins (Hounsell et al, 1985; Hounsell et al, 1989). Core 5 and core 6 (Figure 2) was found to be in fetal mucins in meconium (Hounsell et al, 1985). In addition, core 5 could be detected in rectal adenocarcinomas (Kurosaka et al, 1983) and core 6 in gastric carcinomas. Core 7 is minor component of bovine submaxillary mucins (Chai et al, 1992a; Chai et al, 1992b) and core 8 structure (Figure 2) occurs in small amounts in human bronchial mucin (van Halbeek et al, 1994). In summary, structures of O-glycans are often abnormal and contribute to the phenotype and biology in cancer cells (Brockhausen, 1999).

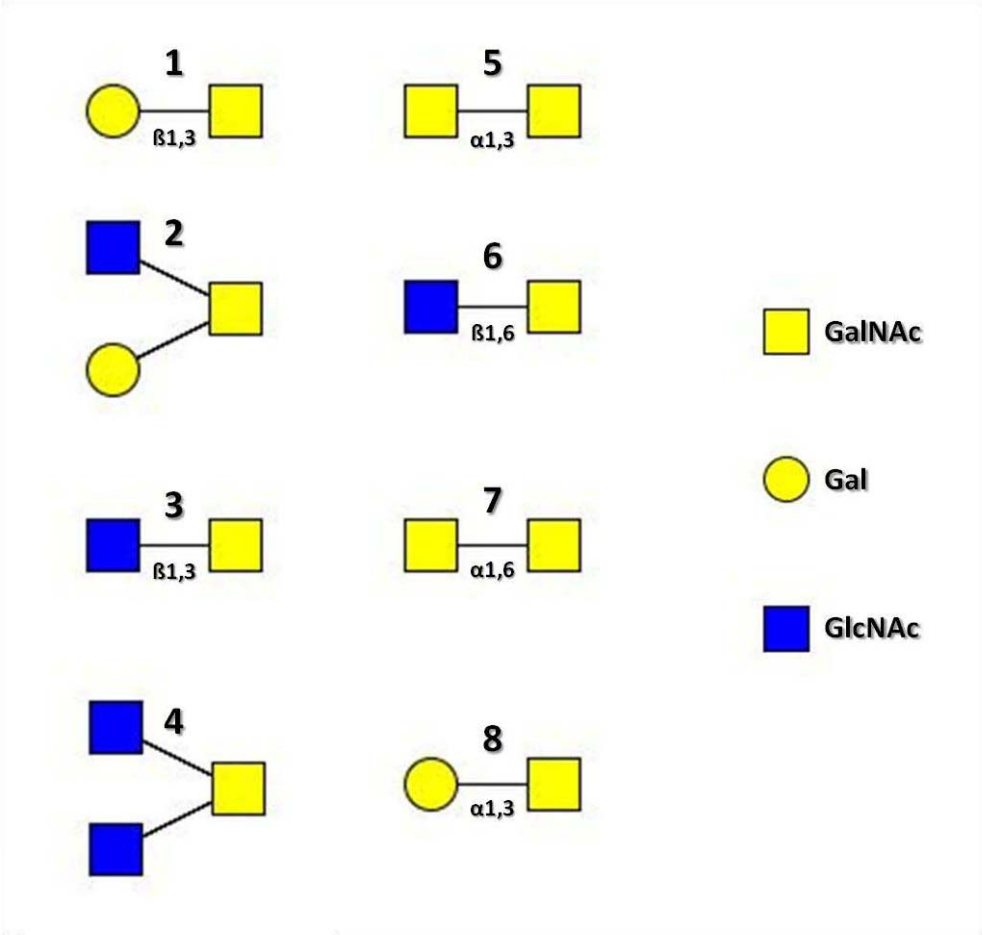

Figure 2: The core structures (1 – 8) of mucin-type O-glycans

1.3 Glycosylation in snails

Snails possess an enzymatic machinery for synthesizing a large number of glycan structures, including typical mammalian, plant, insect or nematode epitopes. In *Lymnea stagnalis* the key enzyme, GlcNAcT-I, for the formation of complex N-glycans was identified. The glycan product resulting from its activity is the prerequisite for the action of GlcNAcT-II, fucosyltransferases (FucT) and xylosyltransferase (XylT) (Mulder et al, 1995a). Furthermore, in several studies it has been shown that *L. stagnalis* contains GlcNAcT-II, XylT (Mulder et al, 1995a), ß1,4-GalNAcT (Mulder et al, 1995b), ß1,3-GalT (Mulder et al, 1991), α1,2-FucT (Mulder et al, 1996), ß1,4-GlcT (van Die et al, 2000), α1,3-FucT catalysing the Lewis^x-unit (Mulder et al, 1996) and an α1,3-FucT catalyzing the transfer of Fuc to the asparigin-linked GlcNAc (van Tetering et al, 1999).

By analyzing neutral N-glycans from *Arion lusitanicus*, *Achatina fulica*, *Limax maximus*, *Cepaea hortensis*, *Planorbius corneus* and *Arianta arbustorum* mainly oligomannosidic and small paucimannosidic structures, often terminated with 3-*O*-methylated Man were detected (Gutternigg et al, 2007). Methylation is a common feature of snail glycans, as first described in *Helix pomatia* that contained 3- or 4-*O*-methylated Gal (Hall et al, 1977; Lommerse et al, 1997). Terminal 3-*O*-methylated Gal residues and 3-*O*-methylated GlcNAc ß1,2-linked to the ß-Man were detected in *Rapana venosa* (Dolashka-Angelova et al, 2003). *P. corneus* and *A. fulica* exhibited traces of large glycan structures, which are terminated by 3-*O*-methylated Gal (Gutternigg et al, 2007). In addition, 4-*O*-methylated Gal isolated from O-glycans was detected in *A. lusitanicus*, *A. fulica*, *A. arbustorum* and *P. corneus* (Stepan et al, 2010).

Truncated glycans of snails often contain ß1,2-linked xylose (Xyl) to the ß-Man residue and/or α-fucosylation, mainly 1,6-linked to the innermost GlcNAc of N-glycans (Gutternigg et al, 2004; Gutternigg et al, 2007). The transfer of Xyl from UDP-Xyl to the core ß-linked Man of N-glycans by ß1,2-XylT is a widespread feature of plant glycoproteins, but very uncommon in the animal kingdom.

O-glycans of 8 different snail species (*A. lusitanicus, A. fulica, B. glabrata, C. hortensis, C. helena, H. pomatia, L. maximus, P. corneus*) were identified and classified into six groups according to the modification of an established trisaccharide core structure, which consisted of GalNAc elongated by two 4-*O*-

methylated Gal residues. The six groups determined were (I) glycans missing one, both methyl groups or one hexose, (II) the core structure containing GalNAc plus two 4-*O*-methylated hexoses, (III) the core linked to additional methylated hexoses, (IV) the core with one or two more unmethylated hexoses, (V) O-glycans containing Fuc residues and (VI) the core containing HexNAc and up to six unmethylated hexoses (Stepan et al, 2012).

In summary, these data motivated us to isolate and characterize the very uncommon enzyme in the animal kingdom, the ß1,2-XylT, as well as two glycosyltransferases responsible for the synthesis of the T-antigen (Galß1-3GalNAcα1-Ser/Thr), namely the ppGalNAcT and the T-synthase, from snail origin.

1.4 Mucin-type O-glycosylation as an important feature for attracting parasites to snail hosts

The host finding process of miracidia (*S. mansoni*) can be divided into at least four different phases, (1) dispersal, (2) microhabitat selection, (3) orientation to the host and (4) behaviour after contact with the host species. Whereas the first two phases are based on environmental factors, the latter phases are mainly stimulated by chemical molecules. Besides small molecular compounds, such as amino acids (MacInnis, 1965), Mg^{++} ions (Stibbs et al, 1976), Ca^{++} concentration (Sponholtz & Short, 1976), peptides (Mason, 1977), fatty acids (MacInnis, 1965), pH (MacInnis, 1965) and ammonia (Mason & Fripp, 1977), it was shown that the miracidia attraction to the host is stimulated by a glycoconjugate (Haberl et al, 1995; Theodoropoulos et al, 2001). Based on out of date methods they used for these experiments, these data should be interpreted carefully, but they represent the only information about molecules necessary for miracidia attraction. The effectiveness of snail conditioned water, containing macromolecules responsible for attracting the miracidia to the snail host (*B. glabrata*), was completely destroyed by alkaline cleavage of *O*-glycosidic linkages. On the other hand, miracidia were still attracted by SCW, when it was treated with endoglycosidase F (Haberl et al, 1995). In summary, these data indicate that O-glycans, but not N-glycans, play an important role for correct host finding.

1.5 Lectins

Based on conserved amino acid sequences within the carbohydrate recognition domain (CRD) Kurt Drickamer published the first classification of animal lectins (Drickamer, 1988). One group required Ca^{++} ions for sugar recognition, so they were named C-type lectins. The other group needed free thiols for stability and the members were termed S-type lectins, which were renamed as galectins nowadays. Lectins that recognize Man-6-P were found to be homologous, but different to the other lectin classes, so these proteins were designated as P-type lectins. The newest group of lectins was identified when it was shown that the immunoglobulin superfamily was able to recognize carbohydrate structures. These lectins were named I-type lectins. A subgroup of the last ones, which specifically binds sialic acid, has been designated Siglecs. Figure 3 shows generic structures of several of these classes of lectins.

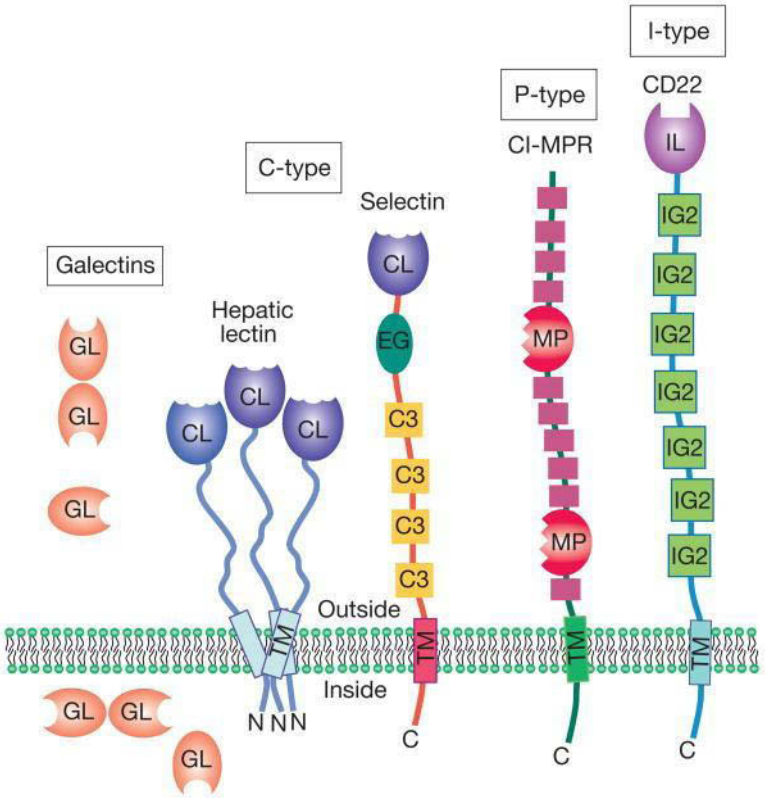

Figure 3: Major types of animal lectins. (taken from Brockhausen et al, 2008)

C-type lectins play an important role for recognition events on cell surfaces and their binding mediate many biological events. These include cell-cell adhesion, serum

glycoprotein turnover and innate immune responses to pathogens (Drickamer & Taylor, 1993).

Lectins contain a homologous CRD that has highly conserved residues contributing to the typical fold of this domain (Weis & Drickamer, 1996). In animals, there are many different proteins containing all or parts of a C-type CRD that is not always restricted to a sugar binding function. Motifs having C-type CRD characteristics are summarized as C-type lectin-like domains (CTLD) (Weis et al, 1998). In *C. elegans* roughly 180 potential CTLDs have been identified (Drickamer & Dodd, 1999). 10% of these are predicted as Ca^{++} dependent sugar binding proteins. Several subfamilies of C-type lectins contain CRDs linked to α-helical domains that have been postulated to form trimers by coiled coil structures, which has been proven for rat and human binding protein (Sheriff et al, 1994; Weis & Drickamer, 1994).

To test sugar binding specificity in earlier times, hemagglutination studies of purified lectin, as described by (Raja et al, 2011) were done. Inhibiting hemagglutination of erythrocytes with different sugar solutions was achieved to get an idea of the sugar binding specificity of lectins. Nowadays, the analysis of the carbohydrate specificity is done on glycan arrays. A great repertoire of glycans are printed on a carrier surface with carbohydrates from mammals or some pathogens, such as from *E. coli, Proteus mirabilis, Pseudomonas aeruginosa, Providencia alcalifaciens, Providencia rustigianii, Providencia stuartii, Shigella boydii*, and *Shigella dysennteriae.*

Lectins from animals and plants are used for different applications in biochemstry and medicine, with detecting of glycan biomarkers in several cancer types as one of the most important one. *Helix pomatia* agglutinin (HPA) is highly specific for GalNAc containing oligosaccharides, protects fertilized eggs from bacteria (Prokop et al, 1968) and is part of the innate immunity system of the snail (Sanchez et al, 2006). This lectin aggregates bacteria, such as group C streptococci (Köhler et al, 1973) and *Listeria monocytogenes* (Patchett et al, 1991). Based on its sugar binding specifity, this snail lectin is used for detecting abnormal glycans in different cancer cells. HPA was applied to sections from primary breast, colon and gastric cancer and bound to tumours of patients with the worst prognosis (Brooks, 2000; Schumacher et al, 2005). It was further proofed that HPA binding is associated with metastases (Schumacher & Adam, 1997).

The fact that invertebrate lectins can be used for several applications in basic and applied research areas was the reason for cloning and recombinant expression of a C-

type lectin from *A. lusitanicus* in this work. In the future sufficient purified protein will be available and used for testing the accurate sugar binding specificity.

2 PURPOSE OF THIS WORK

The aim of this research work was to gain biochemical information about glycosyltransferases from gastropods that synthesize glycan epitopes relevant for various biological functions. Besides mammals, or specific invertebrates, such as *Drosophila melanogaster* and *Caenorhabditis elegans*, there are less data concerning enzymes involved in glycan biosynthesis from gastropods available.

To achieve a basis for future research that study the biological relevance of glycans for host-parasite interactions, in this work three glycosyltransferases should be characterized. Previous studies could show that the mucin-type O-glycans represent the crucial molecules for attracting the human parasite *Schistosoma mansoni* to its intermediate host, the watersnail *Biomphalaria glabrata*.

The most common mucin-type O-glycan in animals, the T-antigen, consists of a disaccharide (Galß1-3GalNAcα1-Ser/Thr) linked to Ser/Thr residues of a polypeptide chain. It is initiated by an enzyme family, called polypeptide *N*-acetylgalactosaminyltransferases (ppGalNAcTs), and elongated by a core-1 ß1,3-galactosyltransferase (core-1 ß1,3-GalT or T-synthase). These enzymes should be isolated, expressed and characterized the first time from *B. glabrata*.

ß1,2-xylosylation of the ß-Man from N-glycans is a typical plant modification, but very uncommon in other animals. However, the intermediate host (*B. glabrata*) and its parasite (*S. mansoni*) possess this extraordinary glycan feature, indicating an important role for interspecies recognition. In addition to the other enzymes mentioned above, a ß1,2-xylosyltransferase (ß1,2-XylT), should be described from snail origin.

3 MATERIALS & METHODS

3.1 Standard procedures

3.1.1 *Dialysis*

To equilibrate a sample to new buffer conditions or to remove salts and other small contaminants, usually dialysis was performed against water or an appropriate buffer. Membranes from regenerated cellulose with a weight cut off of 12000 – 14000 Da and a pore diameter of 25A were used (Servapor).

3.1.2 *Ultrafiltration*

This technique was performed mainly for volume reduction of enzyme containing solutions at 4°C. Stirred ultrafiltration cells (Amicon Bioseparations; Millipore) containing regenerated cellulose filters with a molecular weight cut off of 30000 Da were used.

3.1.3 *Determination of protein content*

Protein concentrations were determined by the Micro-BCA protein assay (Pierce) with bovine serum albumin as the standard.

3.1.4 *SDS-PAGE and Western blot*

Specific sample preparation of cells or tissue is described in respective chapters. Proteins in aqueous solution were precipitated by adding 4 volumes of cold MeOH, incubated for 30min at -80°C followed by a centrifugation step at 14500rpm for

30min (4°C). Resuspended protein samples were diluted 1:1 with sample buffer (Table 1) and boiled at 96°C for 5min. Proteins were separated for approximately 60min at 200V by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) with gels containing 12.5% acrylamide and 1% bisacrylamide, respectively. For Coomassie staining the polyacrylamide gel was fixed in buffer (50% MeOH, 7% HAc in water) for 30min at room temperature prior to incubation in Coomassie solution (Table 1) under weak agitation.

After soaking filters and the nitrocellulose membrane in Blotting buffer (Table 1) the protein transfer was done in a "semi-dry" blotting apparatus for about 35min at 15V. The quality of protein transfer was checked by staining with Ponceau's solution (0.5% (w/v) in 1% HAc). The nitrocellulose sheets were blocked with Blocking buffer (Table 1) either at room temperature for 2h or at 4°C overnight, respectively. Afterwards, the membrane was incubated with primary antibodies or biotinylated lectins diluted to optimal concentrations in Blocking buffer for 1h at room temperature under constant agitation. To remove excess of primary antibodies/lectins the sheets were washed three times with TBS-T buffer (Table 1) followed by the incubation with secondary antibodies or Streptavidin, both conjugated to alkaline phsophatase. Membranes were washed again three times with water, before colour detection was performed using SIGMAFAST™ BCIP®/NBT (Sigma-Aldrich).

Table 1: Reagents for SDS-PAGE and Western Blot

Buffer	Composition
Sample buffer (2x)	31mg DTT 200mg SDS 5ml 0.5M Tris/HCl pH 6.8 2.8ml glycerine (87%) 2.7ml H_2O and bromphenole blue
Coomassie solution	3.5% perchloric acid 0.04% Coomassie Brilliant Blue G-250
10x TBS buffer	1M Tris 1M NaCl pH 7.5
TBS-T buffer	1x TBS + 0.05 Tween
Blotting buffer (1x)	25mM Tris 200mM glycine 20% MetOH in water
Blocking buffer	1x TBS-T + 0.5% BSA

3.1.5 *Dot blot*

Protein samples were spottet onto nitrocellulose membrane and dried for 30min at room temperature. Non-specific sites were blocked by incubating the membrane in Blocking buffer (Table 1) for two hours at room temperature. ß1,2-xylosylated glycans were detected by incubation with anti-horseradish peroxidase (HRP) (produced in mouse; 1:2000) followed by anti-mouse IgG conjugated to alkaline phosphatase (1:2000) for 50min at room temperature. Between the antibody incubations the membrane was washed twice with TBS-T buffer (Table 1) for 20min at room temperature. After antibody incubation the membrane was washed once in TBS-T buffer for 15min, followed by two washing steps in water for 10min. Colour detection was performed using SIGMAFAST™ BCIP®/NBT (Sigma-Aldrich).

3.1.6 *Orcinol test*

Free sugars were detected on TLC plates, which were coated with 100µl orcinol solution (100mg orcinol were dissolved in 50ml 20% H_2SO_4) and dried using a blow dryer. 1µl of sample was spotted on the plate, dried and incubated for 5min at 100°C. Samples containing sugars developed colour.

3.1.7 *Microsome preparation*

About 50mg of frozen snail tissue (*A. lusitanicus*; *A. fulica*) was homogenized in buffer (5mM imidazole, 250mM sucrose pH 7.3) using UltraTurrax (3x at maximum speed for 30sec). The homogenate was centrifuged at 6000rpm for 25min (4°C) to remove cell depris. The supernatant was used to pellet microsomes at 19000rpm for 20min (4°C) in an ultracentrifuge. The pellet was dissolved in 100µl buffer (5mM imidazole, 250mM sucrose 1% Triton X-100 pH 7.3) for downstream processes, such as enzyme activity assays.

3.1.8 *GnGn glycan isolation from fibrin*

3g fibrin were suspended in 100ml buffer (0.15M Tris/HCl pH 7.8 + 1mM $CaCl_2$ + 0,02% NaN_3) and incubated for 25min at 100°C. 100mg Pronase was suspended in 1.5ml of the same buffer, mixed with the fibrin solution and incubated overnight at 37°C. Unsoluble material was removed by centrifugation at 5000rpm for 15min (4°C) and the supernatant was reduced to a volume of about 2ml using a rotavapor. The sample was loaded to a Sephadex G25 fine column (1.5 x 120cm) and separated with 1% HAc. 80 drops were collected per fraction, which were tested for the presence of peptides and carbohydrates by UV absorption (280nm) and orcinol test (chapter 3.1.6), respectively. Fractions containing glycopeptides were pooled and lyophilized. For desialylation 50ml 0.05M H_2SO_4 were added to the sample, incubated for 1h at 80°C and loaded again on the Sephadex G25 column.

To remove N-glycans from the peptide backbone, dry glycopeptides were resuspended in 1ml of PNGase A buffer (50mM citrate/phosphate buffer pH 5.0), 6µl PNGase A (*N*-glycopeptidase A from almond; 5mU/100µl) were added and incubated overnight at 37°C. Salts and the enzyme were removed using the Sephadex G25 column as described above. Glycan containing fractions were pooled and the pH decreased to 2.0 by adding HAc. The sample was loaded to an ion exchange column (AG50WX2) and glycans were eluted with 2% HAc in water. Remaining glycopeptides were eluted with buffer (0.4M NH_4Ac pH 6.0) and treated again with PNGase A. Free glycans were pooled and lyophilized overnight. Degalactosylation was performed by adding buffer (50mM NaCitrat pH 4.6 + 0,04% NaN_3) and 1.5µl galactosidase (Sigma; 5U/mg protein) to the lyophilized sample overnight at 37°C. The sample was loaded again to the Sephadex G25 column under same conditions as described above, lyophilized and resuspended in 500µl water. Quality control of isolated GnGn was performed by HPLC (chapter 3.10.2).

3.2 Cell culture

3.2.1 *Escherichia coli*

JM109 (Promega), NEB10-beta Competent (High Efficiency; New England Bioloabs) and BL21 (New England Biolabs) cells were grown in Lysogeny broth (LB) medium (1% Bacto-tryptone 0.5% Bacto-yeast extract 0.5% NaCl) at 37°C and 220rpm, or on LB agar plates (1.5% agar) at 37°C. After transformation of recombinant plasmids the adequate antibiotic was added to the medium (100 µg/ml Ampicillin, 25µg/ml Zeocin, 50µg/ml Kanamycin; 50µg/ml Chloramphenicol).

3.2.2 *Yeast cells (Pichia pastoris)*

X-33 (Invitrogen) cells were grown in YPD medium (1% yeast extract, 2% peptone, 2% glucose) at 30°C and 200rpm, or on YPD agar plates (1.5% agar) at 30°C. Following transformation of linearized, recombinant plasmid DNA (pGAPZ B), 25µg/ml Zeocin was added to the medium.

3.2.3 *Insect cells*

*Spodoptera frugiperda Sf*9 (ATCC CRL-1711) (Summers and Smith 1987) and *Spodoptera frugiperda Sf*21 (Vaughn et al, 1977) cells were cultivated in IPL-41 medium (SAFC Biosciences) containing yeast extract, a lipid mixture supplemented with 3% fetal calf serum (FCS), at 27°C using T-flasks. SfSWT-1 Mimic^{TM} insect cells (Invitrogen) (Hollister et al, 2002) cells were grown under same conditions, but with 10% FCS. In the case of *Trichoplusia ni* BTI-TN5B1-4 “Hi5” cells (ATCCCRL-10859) (Wickham & Nemerow, 1993) and *Ascalapha odorata* cell line, BTI-*Tnao*38 (Hashimoto et al, 2010), no FCS was present in the medium.

3.2.4 *Biomphalaria glabrata embryonic (Bge) cells*

90ml basic medium of Bge cells were prepared by mixing 20ml MilliQ water, 22ml Schneider's Drosophila Medium (Sigma-Aldrich), 2ml lactalbumin hydrolysate (sterile filtered; 225mg/ml), 1ml Gal (sterile filtered; 130mg/ml), 40µl gentamycin (Gibco; 50mg/ml), and ~162µl phenolred (0,5% solution). The pH was adjusted to 7.0 with 1M NaOH (~140µl) and water was added to a final volume of 90ml. For preparing the complete medium, 10ml of heat inactivated FCS (Gibco; FBS South American heat inactivated, Cat. Nr. 10500064) was added to 90ml basic medium. The cells were grown at 26°C and splitted 1:2 or 1:3 every three/four days. Detachment was possible by firm tapping. They needed at least 24h to reattach completely.

3.3 Creation of Cryostocks

3.3.1 *E. coli*

An overnight culture was mixed with glycerol to a final concentration of 25% and stored at -80°C in 500µl aliquots.

3.3.2 *Yeast cells*

An overnight culture was mixed with glycerol to a final concentration of 15% and stored at -80°C in 500µl aliquots.

3.3.3 *Bge cells*

Cells were centrifuged at 200xg for 10min, the supernatant discarded and 900µl of cold medium (90% FCS, 10% DMSO) was added dropwise to the cell pellet. The cells were transferred to cryovials and incubated for 4h at 4°C followed by 24h at -80°C. When they were completely frozen, the vials were put quickly to liquid nitrogen. After a couple of days viability was checked by thawing an aliquot.
The thawing procedure was done as follows: Frozen cells were put in a water bath at 35° – 37°C and incubated until 80% were thawed. There should be still a small piece of ice in the tube. Then 1ml of Bge basic medium at room temperature was added. The cells were transferred to a 15ml falcon tube and centrifuged at 700xg for 5min. The pellet was washed once with 11ml basic medium and finally resuspended in 5ml complete medium (chapter 3.2.3). This cell suspension was transferred to a T25 flask and incubated for one week at 26°C without changing the medium. Up to four weeks or until the cells were confluent, respectively, the complete medium was changed. After approximately four weeks the cells can be splitted as described in chapter 3.2.3.

3.4 Cloning

3.4.1 *Gel electrophoresis*

Gel electrophoresis was performed with 1% agarose gels at 120V for about 30min. 100V for 60min was done for DNA purification procedures from agarose gels. The running time was dependent on DNA size.

3.4.2 *DNA purification*

DNA extraction from agarose gels or purification from PCR/restriction mixtures was performed using NucleoSpin® Gel and PCR Clean-up (Machery Nagel). PCR/restriction reactions were adjusted to 100µl with water and 2 volumes of Buffer NTI were added. In the case of agarose gels containing the desired DNA, 200µl Buffer NTI were added per 100mg gel and dissolved for at least 10min at 50°C. Up to 700µl of dissolved DNA was loaded to the "NucleoSpin® Gel and PCR Clean-up Column" and centrifuged for 30sec at 11000xg. Bound DNA was washed with 700µl Buffer NT3 (30sec at 11000xg). The column was dried completely (1min at 11000xg), before the DNA was eluted in 15 – 30µl water (1min at 11000xg).

3.4.3 *Restriction*

The restriction mixture comprised of at least 1µg DNA, the appropriate buffer, restriction enzyme(s) and 100µg/ml BSA, if necessary, in a total volume of 10-50µl. The restriction was performed at 37°C for 1 – 3h. All enzymes and buffers were used from New England Biolabs.

3.4.4 *PCR*

Standard PCR reactions contained at least 1ng DNA as the template, the buffer, 1.5mM $MgCl_2$, 0.2mM dNTP's each, ~0.3µM primers and 1.5U DNA Polymerase in a total volume of 20-30µl. The PCR program started usually with an initial denaturation step at 94°C for 4min, followed by 30 cycles (94°C 30sec, 50 - 65°C 30sec, 72°C 1min) and a final incubation time at 72°C for 5min. The annealing temperature/time, as well as the elongation time varied in terms of primer features and the fragment length to amplify, respectively.

3.4.5 *DNA precipitation*

DNA resuspended in water was mixed with 2.8x volumes of EtOH and glycogen (100ng/µl), incubated at least for 1h at -80°C and centrifuged at 13500rpm for 30min (4°C). The supernatant was discarded carefully and the DNA pellet dried at room temperature for exactly 10min. Finally the DNA pellet was resuspended in 10 – 20µl water and its amount or purity was measured using NanoDrop.

3.4.6 *Ligation*

Standard ligation reactions contained about 100ng vector DNA, 5-7x molar excess of insert DNA, T4 buffer and T4 DNA ligase in a total volume of 10µl. The ligase and the buffer were used from New England Biolabs. The ligation reaction was performed at 14°C overnight.

3.4.7 *Plasmid isolation*

Plasmids were purified from *E. coli* cells using NucleoSpin® Plasmid Quick Pure (Macherey Nagel). Cells of a 5ml overnight culture (ONC) were harvested (30sec at 11000xg), resuspended in 250µl Buffer A1 followed by adding 250µl Buffer A2.

The cell suspension was inverted 6-8 times and incubated for 5min at room temperature. After neutralizing the reaction with 300µl of Buffer A3, the sample was clarified by a centrifugation step (5min for 11000xg). The clear supernatant was loaded on a column, centrifuged (1min at 11000xg) and washed with 700µl Buffer A4. The column was dried by centrifugation (2min at 11000xg) and bound plasmid DNA was eluted with 50µl water (1min at 11000xg).

3.4.8 *Sequencing*

All samples were sequenced at Eurofins MWG Operon.

3.5 Transformation of *E. coli*

3.5.1 *Electroporation*

10µl of the ligation mixture was precipitated with EtOH (chapter 3.4.5). The dried DNA pellet was resuspended in 10µl of water and 1/10 was used for the electroporation of 50µl competent JM109 cells. The electroporation was performed at 2500V, 25µF and 200Ω.

3.5.2 *Chemical transformation*

The frozen chemical competent cells (NEB10ß Competent (High Efficiency; New England Biolabs) & BL21 (New England Biolabs)) were thawed on ice for 10min. Up to 100ng of plasmid DNA was added, carefully flicked and incubated for 30min on ice. The following heat shock was performed at 42°C for exactly 30sec before the transformed cells were cooled down for 5min on ice. Furthermore, 950µl SOC medium (Peptone 20g/l; Yeast extract 5g/l; NaCl 10mM; KCl 3mM; $MgCl_2$ 10mM; $MgSO_4$ 10mM; glucose 20mM) at room temperature was mixed with the cells and incubated for 60min at 37°C and 250rpm. Aliquots of transformed cells were spread on LB agar plates containing the adequate antibiotic and incubated overnight at 37°C.

3.5.3 *Colony PCR*

Transformed *E. coli* colonies, which had been grown overnight at 37°C, were resuspended in 50µl of water, boiled for 8min at 96°C and centrifuged for 30sec at 11000rpm. 5µl supernatant was used as the template for PCR under standard conditions (chapter 3.4.4) by using primers flanking the multiple cloning site.

3.6 Transformation of *P. pastoris*

3.6.1 *Creation of electrocompetent yeast cells (X-33)*

Based on a 5ml ONC, 500ml YPD medium was inoculated to OD_{600} 0.2 – 0.7. The cells were grown to OD_{600} 1.4 – 2.0 and harvested at 2000xg for 10min at 4°C. The pellet was resuspended in 100ml YPD containing 20mM HEPES pH 8.0 and 2.5% DTT (v/v). The cell suspension was incubated shortly at 30°C under constant agitation before 400ml ice cold water was added. From now on all the steps were done at 4°C. After centrifugation at 2000xg for 10min, the cell pellet was washed first with 250ml cold 1mM HEPES pH 6.0 and then with 200ml cold 1M sorbitol. The final pellet was further resuspended in 500µl 1M sorbitol. 80µl aliquots were frozen at -80°C for future electroporation.

3.6.2 *Transformation of electrocompetent yeast cells (X-33)*

20µl of linearized recombinant plasmid DNA (pGAPZ B) was mixed with 80µl aliquots of electrocompetent yeast cells. The electroporation was performed at 2kV, 25µF and 200Ω. 900µl YPD medium at room temperature was added to the cells and incubated for 2h at 30°C without any agitation. Aliquots were spread on agar plates and incubated for 2 days at 30°C.

3.6.3 *Colony PCR*

Transformed yeast cells were resuspended in water and sonicated (Duty Cycle = 30; Output Control = 3; 3min) before the cell suspension was boiled for 20min at 96°C. Afterwards the sample was centrifuged for 30sec at 800rpm. 1µl of the supernatant was used as the template for the control PCR with AOX3' (5'GCAAATGGCATTCTGACATCC3') and AOX5' (5'GACTGGTTCCA ATTGACAAGC3') primers. As a positive control the GADPH primers (forward:

5’ATGGCTATCACTGTCGGTATTAACG3’) and reverse (5’TTAAGCCTTAG CAACGTGTTGC3’) were used.

3.7 Expression in different host cells (*E. coli*, *P. pastoris* and insect cells)

3.7.1 *E. coli (BL21)*

5ml LB media containing 50µg/ml Kanamycin/Chloramphenicol was inoculated with single BL21 colonies containing recombinant pET26b constructs (Figure 4) and incubated overnight at 37°C and 220rpm. The next day 50ml LB media (50µg/ml Kanamycin/Chloramphenicol) was inoculated to a OD_{600} of 0.1 and incubated until OD_{600} of 0.6 – 0.8. By adding IPTG in a final concentration of 0.1mM the recombinant expression was induced at 37°C. After 1h, 3h and 24h the same biomass was taken for testing the optimal expression via Western Blots (chapter 3.1.4). Cell lysis was performed by sonication (Duty Cycle = 30; Output Control = 3; 3min). Recombinant expression was checked in the cell pellet and the supernatant resulting from centrifugation of lysed cells for 30min at 14500rpm.

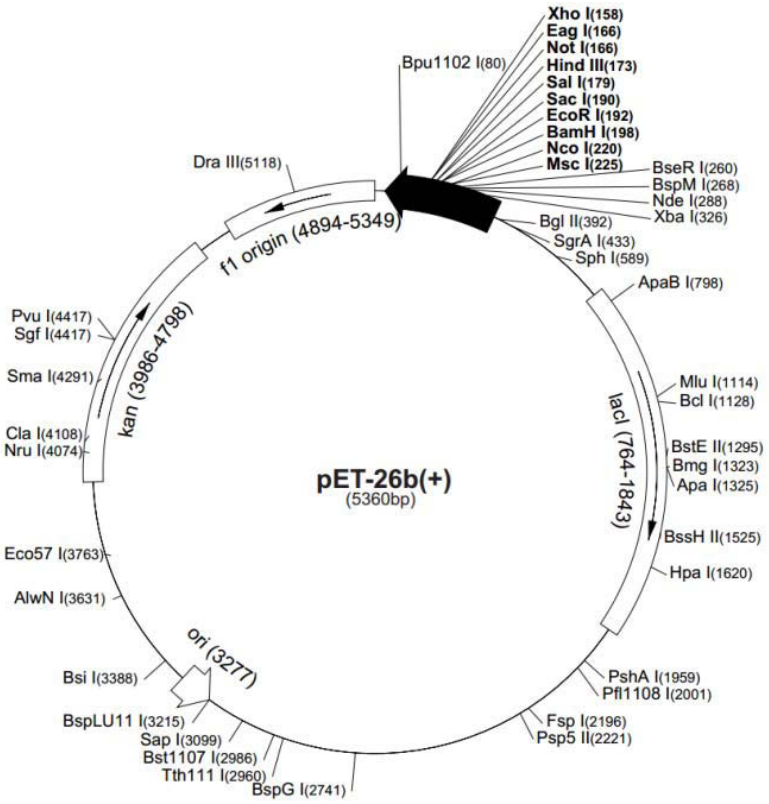

Figure 4: Vector map of pET26b. (taken from: http://www.merckmillipore.de/life-science-research/vector-table-novagen-pet-vector-table/c_HdSb.s1

O77QAAAEhPqsLdcab?PortalCatalogID= merck4biosciences&CountryName =Austria; 5.9.2013)

3.7.2 *P. pastoris*

Positive clones checked by Colony PCR (chapter 3.6.3) containing recombinant pPICZα constructs (Figure 5) were first grown in 4ml YPD medium + 50µg/ml Zeocin overnight at 30°C and 200rpm. The cell pellet (1000xg for 5min) was resuspended in 25ml YPD medium + 0.5% MetOH (v/v) and incubated at 24°C for three days. Every 24h 0.5% MeOH (v/v) was added to the yeast culture. Recombinant expression was checked by Western blot.

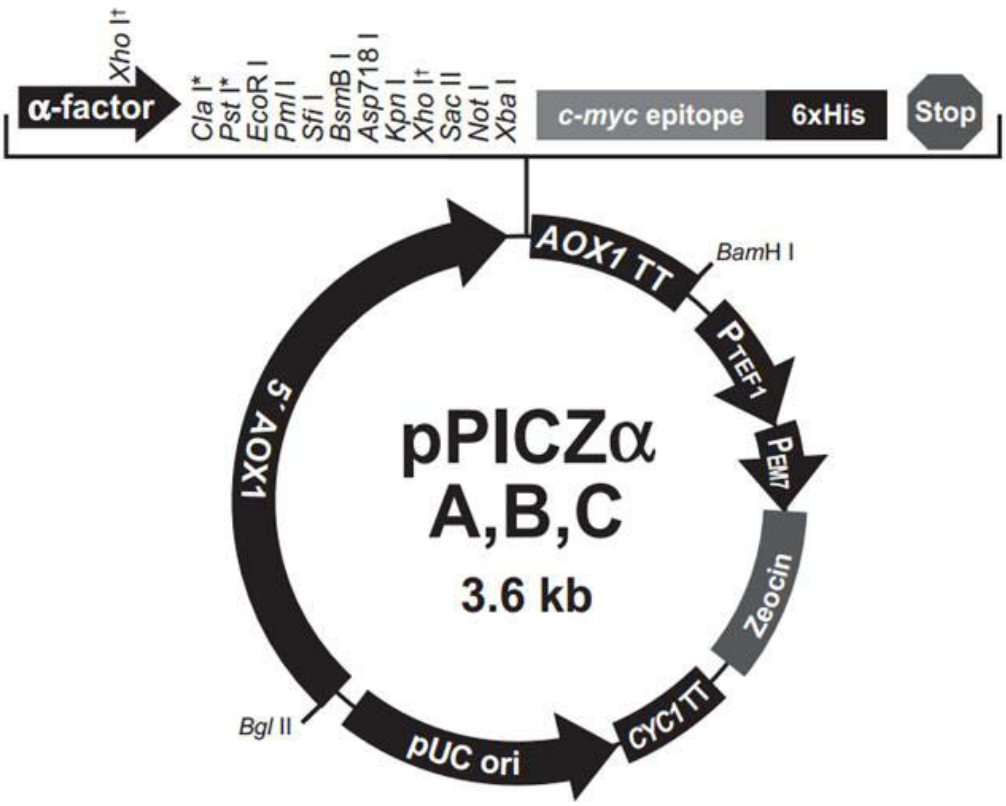

Figure 5: Vector map of pPICZα A. (taken from http://www.lifetechnologies.com/order/catalog/product/V19520; 28.10.13). Vector was modified by Dr. Martin Dragosits (insertion of an N-terminal 6x His/Flag tag, respectively)

3.7.3 *Insect cells*

Based on the Co-Transfection (chapter 3.8.1) a three step upscaling procedure was performed to yield a high titer of baculovirus. Finally $1.5x10^7$ *Sf*9 logarithmically growing insect cells were infected with the baculovirus (MOI ~ 5). The supernatant or the insect cells, respectively, were harvested after 3 days for testing recombinant protein expression.

3.8 Creation of recombinant baculovirus for expression in insect cells

3.8.1 *Co-Transfection*

*Sf*9 insect cells were splitted (1:8 – 1:10) and grown for 3 days under standard conditions (chapter 3.2.3). One day before the cells were co-transfected, a 1:2 dilution was performed. $8x10^5$ cells/well (6-well plate) were incubated for 30min at room temperature allowing cell adhesion. To remove the fetal calf serum completely, the cells were washed three times with IPL-41. Finally 1ml of IPL-41 was added to the insect cells. For the Co-transfection with pBacPAK8 constructs (Figure 6) two solutions (A + B) were prepared. Solution A contained 9% Cellfectin II (Invitrogen) in IPL-41 (v/v) in a total volume of 100µl and solution B comprised 60ng linearized BaculoGold DNA (BD Biosciences), 600ng of recombinant pBacPAK8 plasmid and IPL-41 in a total volume of 100µl. Both solutions were mixed and incubated for 20min at room temperature. The 200µl mixture was added to the cells, incubated for 5h at 27°C and replaced by culture medium containing 50µg/ml Penicillin/Streptomycin. Co-transfected cells were incubated for 5-6 days under standard conditions (chapter 3.2.3).

Solution A used for the Co-transfection with pVT-Bac-His – constructs (Figure 7) contained 3% Lipofectin (Invitrogen) in IPL-41 (v/v) in a total volume of 1ml. Solution B was generated by mixing 300ng linearized BaculoGold DNA (BD Biosciences), 2µg of recombinant pVT-Bac-His constructs and 1ml IPL-41. The mixture of A and B was incubated for 30min at room temperature prior to adding it to the cells. The following procedure was similar as described above for pBacPAK8 constructs.

In the case of a third vector, recombinant baculovirus DNA was produced in vitro by mixing 300ng of recombinant pENTR1A constructs (Figure 8), 300ng isolated and SceI restricted baculovirus genome, 50mM TE buffer and Clonase (Invitrogen) enzyme in a total volume of 20µl. The mixture was incubated overnight at room temperature. Solution B contained 10µl recombination mixture in a total volume of 100µl IPL-41, whereas solution A, as well as the following transfection procedure, was same as described for pBacPAK8 constructs.

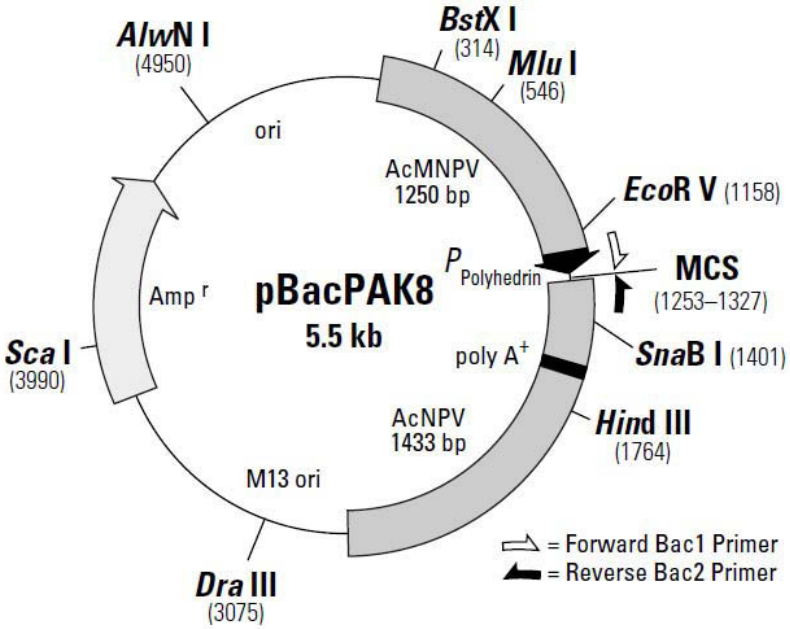

Figure 6: Vector map of pBacPAK8. (taken from Clontech Laboratories; www.clontech.com; 28.10.13)

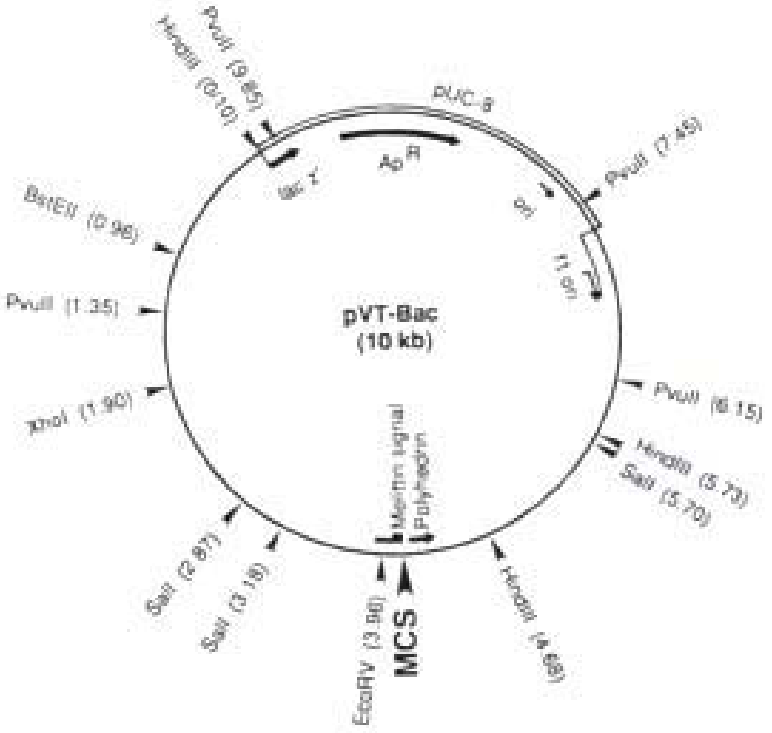

Figure 7: Vector map of pVT-Bac. (taken from Tessier et al, 1991). pVT-Bac-His-1 is a modified version of pVT-Bac. In contrast to the parent plasmid, the multiple cloning site of pVT-Bac-His-1 contains the coding sequence of a N-terminal 6 x His tag

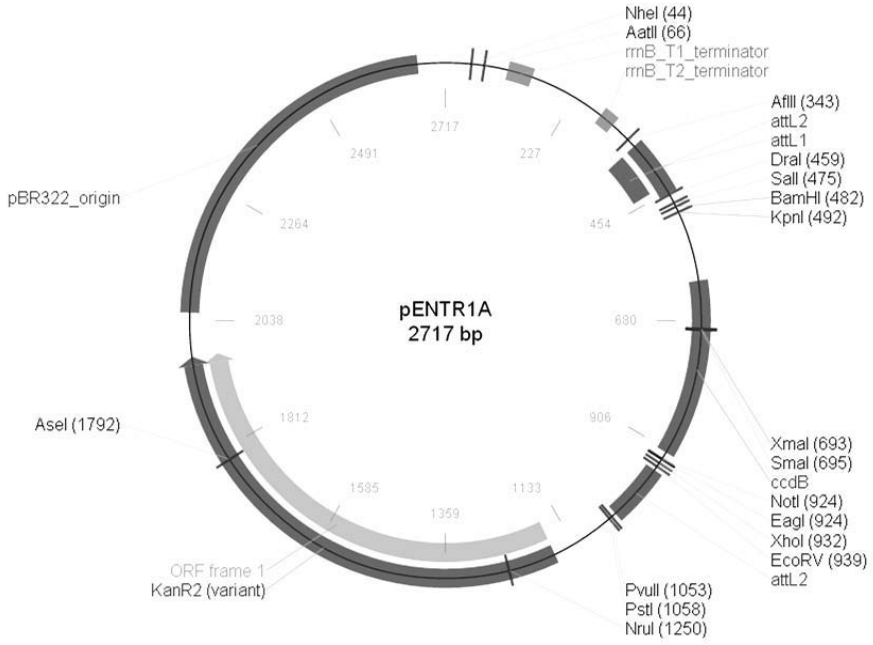

Figure 8: Vector map of pENTR1A. (taken from http://www.addgene.org/vector-database/2525/; 5.9.2013)

3.8.2 *Virus DNA isolation*

$1x10^6$ *Sf*9 cells were treated the same way as described in chapter 3.4.2 until the neutralization step (addition of Buffer A3). To avoid damage of the Baculovirus genome, the DNA was not bound to the column, but precipitated with EtOH (chapter 3.4.5). The isolated DNA was used as template for testing the correct insertion within the Baculovirus genome by PCR.

3.8.3 *Plaque assay*

*Sf*9 cells were grown for 3 days under standard conditions (chapter 3.2.3) and splittet 1:2 the day before the plaque assay was performed. $2x10^6$ cells were spread per plate and incubated for approximately 1h at room temperature. After removing the supernatant, 500µl of virus dilutions (diluted until 10^{-6}) were added to adherent cells

and incubated for 1h. For spreading the virus solution homogeneously, every 15min the plates were shaked gently. After the supernatant was removed, 4ml 38°C temperate agar solution (solution A and B was mixed shortly before overlaying the cells) per plate was added to the cells and incubated for 15min at room temperature. Solution A comprised 20% FCS within IPL-41 + 50µg/ml Penicillin/Streptomycin in a total volume of 20ml. Solution B contained 2% (w/v) low melt agarose in 20ml IPL-41 media, which was boiled in a microwave for ~2min at 180W. The plates were further sealed in a plastic bag and incubated at 27°C for 6 days. Then 500µl of 1:5 diluted "Thiazolyl Blue Tetrazolium Bromide dye" (Sigma Aldrich) in IPL-41 was added to the agar and incubated for 45min at room temperature for distinguishing dead and living cells.

3.9 Synthesis of double-stranded cDNA and cDNA library construction

Total RNA was isolated from *A. lusitanicus*, *A. fulica*, *B. glabrata* and Bge cells by RNeasy Mini Kit (Qiagen). Polyadenylated RNA was purified from total RNA by MicroPoly(A)Purist^{TM} Kit (Ambion). Approximately 100ng polyadenylated RNA were used as a template for first strand cDNA synthesis by In-Fusion® SMARTer™ Directional cDNA Library Construction kit (Takara). For the construction of libraries in *E. coli*, 200ng cDNA were recombined with 50ng linearized pSMART2IFD, a vector included in the kit. Before the recombination was done the cDNA was precipitated with EtOH and resuspended in water. 1µl of the resulting cDNA library was transformed chemically into NEB 10-ß Competent *E. coli* cells and incubated overnight at 37°C. Colony PCR (chapter 3.5.3) was performed by using the primer set 5'TCACACAGGAAACAGCTATGA3' (forward) and 5'CCTCTTCGCTATTACGCCAGC3' (reverse).

The yeast vector pGAPZ B (Figure 9) was amplified by inverse PCR, prior to the recombination with synthesized cDNA using "In-Fusion SMARTer cDNA Library Construction Kit". For the introduction of specific sequences required for the recombination procedure, the forward primer 5'TTGATACCACTGCTTCCGCCAGCTTTCTAGAAC3' and the reverse primer 5'TTGATACCACTGCTTCCTCGTTTCGAAATAGTTG3' were used. The PCR program started with 98°C for 30sec, followed by 30 cycles of 98°C for 10sec, 60°C for 30sec, 72°C for 1min 40sec, and was finished by 72°C for 5min. 300ng of purified and EtOH precipitated vector was recombined with 750ng cDNA and transformed chemically into NEB10-ß cells. Recombinant plasmids were isolated from *E. coli* cells and transformed into X-33 yeast cells (chapter 3.6.2). TOY3, a mutant yeast strain that is possible to convert UDP-glucose to UDP-xylose (Oka & Jigami, 2006), was not transformed by the snail cDNA library.

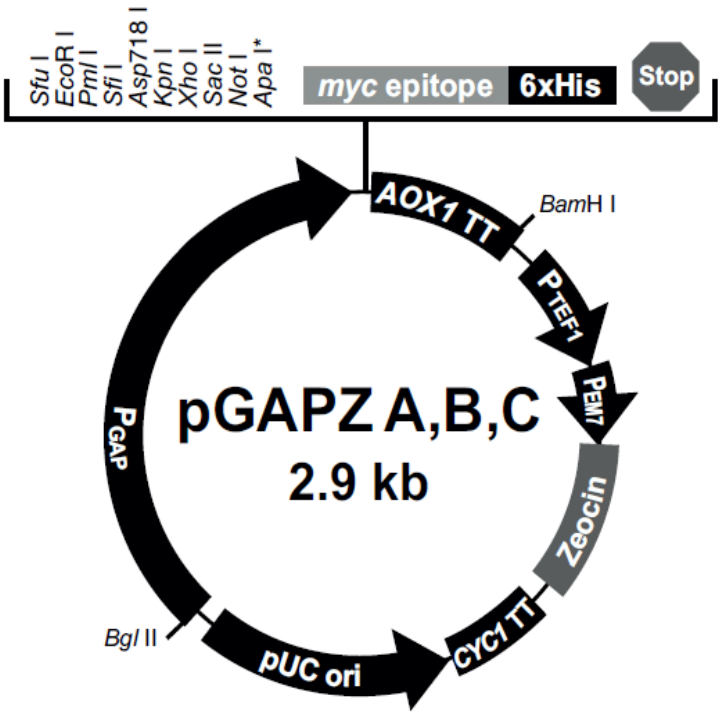

Figure 9: Vector map of pGAPZ B. (taken from www.invitrogen.com; 20.10.13)

pBacPAK8 vector (Figure 6) was amplified by inverse PCR for introducing flanking regions necessary from the recombination with the synthesized cDNA. The PCR started with of an initial denaturation at 98°C for 30sec followed by 35 cycles (98°C for 10sec, 61°C for 30sec 72°C for 2min) and a final extension at 72°C for 5min. The amplification was performed using isolated pBacPAK8 vector as template, 5´TTGATACCACTGCTTGGCCTCGAGT TCGAATCTAG3' and 5'TTGATACCACTGCTTTGCAGGGATCCGTATTTATAG3' as the primer set. Before recombination of cDNA and pBacPAK8 vector, the vector was purified and precipitated with EtOH (chapter 3.4.5). 100ng of synthesized pBacPAK8 vector was recombined with 180ng cDNA from *A. lusitanicus* (In-Fusion™ SMARTer™ cDNA Library Construction Kit; Clontech) for creating an expression cDNA library in insect cells. Isolated pBacPAK8 plasmids were co-transfected with linearized BaculoGold DNA as described in chapter 3.8.1.

3.10 Analytical tools in biochemistry and molecular biology

3.10.1 *2-aminopyridine (2-PA) labelling of N-glycans*

The sample was dissolved in 80µl 2-PA working solution (Table 2) and boiled for 13min. After adding 4µl $NaCNBH_3$ solution (Table 2) the sample was incubated at 90°C overnight. The excess of 2-PA was removed using Sephadex G15 and glycan containing fractions were pooled and lyophilized. Labelled N-glycans were resuspended in water and analyzed by HPLC.

Table 2: Reagents used for 2-aminopyridine labelling of N-glycans

2-PA stock solution	2-PA working solution	$NaCNBH_3$ solution
1g 2-PA in conc. HCl	Diluted stock solution: 1:3 pH should be 6.2	10mg $NaCNBH_3$ in 20µl 2-PA working solution + 30µl water, freshly prepared

3.10.2 *N-glycan analysis by HPLC*

3.10.2.1 Reversed phase

2-aminopyridine (PA)-labelled N-glycans were separated using a reversed phase C18 column (ODS Hypersil, 5µm, 250x4mm, Thermo Scientific) at a flow rate of 1.5ml/min. Fluorometric detection was performed at 320nm excitation and 400nm emission. Solvent A was composed of 0,1M NH_4Ac pH 4.0 in water. Solvent B contained 30% MeOH in water. The following gradient was performed:

Time (min)	%B
0	0
30	30
31	0
35	0

3.10.2.2 Normal phase

2-PA-labelled N-glycans were separated using a normal phase column (Palpak, 5µm, 250x4mm) at a flow rate of 1.0ml/min. Fluorometric detection was performed at 310nm excitation and 380nm emission. Solvent A was composed of 50% AcCN and 50% of "3% HAc pH 7.3 adjusted with triethylamine, 10%AcCN". Solvent B was pure AcCN. The following gradient was performed:

Time (min)	%B
0	45
5	45
45	0
46	0
53	45

3.10.3 *Monosaccharide analysis by HPLC*

Proteins were precipitated by adding four volumes of cold MeOH to samples dissolved in aqueous solution. To remove polysaccharide contamination of animal and fungi samples, a further precipitation was performed. Monosaccharides from glycoproteins were hydrolyzed with 4M trifluoroacetic acid (TFA) by an incubation at 115°C for 2 hours. After drying the samples using a centrifugal evaporator the samples were dried two times in the presence of 30% and once 100% of MeOH. The labeling procedure with anthranilic acid (AA) was done according to (Stepan & Staudacher, 2011). Dry monosaccharides were dissolved in 5µl of sodium acetate trihydrate solution (80mg/ml) before adding 10µl of AA reagent solution (3mg of AA and 30mg of sodium cyanoborohydride in 1ml of 2% [w/v] boric acid in methanol). This mixture was incubated for 60min at 80°C and diluted 1:500 prior to HPLC analysis. Monosaccharides were separated using a reversed phase C18 column (ODS Hypersil, 5µm, 250x4mm, Thermo Scientific) at a flow rate of 1ml/min. Fluorometric detection at 360nm excitation and 425nm emission was performed. Solvent A was composed of 1.0% tetrahydrofuran (v/v), 0.5% phosphoric acid (v/v), and 0.2% 1-butylamine (v/v) in water. Solvent B was a mixture of 50% solvent A and 50% acetonitrile. The following gradient was performed:

Time	%B
0	5
17	25
23	100
29	100
30	5
35	5

Labelling of monosaccharides with 1-phenyl-3-methyl-5-pyrazolone (PMP) was done according to (Stepan & Staudacher, 2011). Acid hydrolysis and the following drying procedure was exactly done as described above. Dried monosaccharides were dissolved in 25µl of 0.5M PMP in methanol. Then 15µl of 0.5M NaOH and 10µl of water was added to the sample and incubated for 120min at 70°C. The neutralization step was performed by adding 20µl of 0.5M HCl. The mixture was extracted 3 - 5 times with diethylether by vigorous mixing followed by a centrifugation at 3000xg for 1min. The organic layer was removed carefully. The final aqueous layer was dried and then redissolved in 1ml of water. HPLC analysis was carried out on a reversed phase C18 column (ODS Hypersil, 5µm, 250x4mm, Thermo Scientific) using a flow rate of 1.5ml/min. Solvent A contained 8% acetonitrile in 0.1M ammonium acetate buffer (pH 5.5), and solvent B was a mixture of 30% acetonitrile in 0.1M ammonium acetate buffer (pH 5.5). Detection was carried out using an UV detector at 245nm. The following gradient was performed:

Time	%B
0	45
10	50
26	85
26.5	85
27	45
32	45

3.10.4 *MALDI-TOF-MS*

MALDI-TOF MS analysis was carried out on an Autoflex Speed MALDI-TOF (Bruker Daltonics, Germany) equipped with a 1,000 Hz Smartbeam.II laser in positivemode using α-cyano-4-hydroxycinnamic acid as matrix. Spectra were processed with the manufacturer's software (Bruker Flexanalysis 3.3.80).

3.10.5 *ETD-MS/MS measurement of glycosylated peptides*

Prior to MS-measurement, enzymatic reaction solution was purified using C-18 SPE cartridges from Thermo Scientific. Briefly, the tubes were equilibrated with MeOH, 60% AcCN and 2 times 0.3% formate buffered to pH 3.0 with ammonia, 500µl each. Sample was applied, washed twice with formate buffer and eluted with 60% AcCN. Peptides were vacuum-dried and redissolved in ddH_2O.

For Muc5Ac labelling with TMT126 mass tag (Thermo Scientific), the peptide was taken up in 50µl 0.1 M triethylammonium bicarbonate buffer (Sigma Aldrich) and spiked with 20µl label reagent. After 1h incubation at RT, the reaction was stopped by adding 8µl 5% hydroxylamine (Sigma Aldrich). Labelled peptides were subsequently C-18 SPE purified, as described above.

Glycosylated peptides were subjected to ETD-MSMS via direct infusion (flow rate 2µl/min) on a Bruker amaZon ion trap. Specific MS-settings were: ICC target 200.000, max accu time 100ms, isolation width 4m/z. ETD reagent transfer time (10-100ms) and target mass (700-950m/z) were adopted specifically for each peptide. Analysis of results was done with Bruker's Data analysis 4.0 and Bio Tools 3.2. Briefly, an average mass spectrum was generated using SNAP peak finder algorithm and exported to Bio Tools. The measured ETD fragment ions were compared to theoretical fragment ion masses originating from every possible glycopeptide form. By comparing confirmed fragment ions for each possible form, real glycosylation site(s) could be elucidated.

3.10.6 *FACS*

The infection of insect cells was done as described in chapter 3.7.3. ~10^6 infected insect cells were harvested (600xg for 10min), washed with 1ml phosphate buffered saline (PBS; 137mM NaCl; 2.7mM KCl; 10mM Na_2HPO_4; 2mM KH_2PO_4) and finally resuspended in 499µl PBS/0.5% FCS. Biotinylated lectins (Vector Laboratories) or primary antibodies (anti-HRP antibody) were diluted 1:500 within the cell suspension and incubated for 1h at room temperature under constant agitation. After 1 – 3 washing steps with PBS (600xg for 10min) the cells were resuspended in 499µl PBS/0.5% FCS. Anti-rabbit IgG conjugated to PE or Streptavidin-PE, respectively, was diluted 1:500 and incubated for 20min in the dark under constant agitation. Labeled cells were washed, centrifuged at 600xg for 10min and resuspended in 500µl PBS.

3.10.7 *MACS*

The infection of insect cells was done as described in chapter 3.7.3. ~5x10^6 insect cells were harvested (300xg for 10min), washed with PBS/0.5% BSA and resuspended in 1ml of the same buffer. Anti-HRP antibody (produced in rabbit) was diluted 1:1000 and incubated for 1h at 4°C under constant shaking. Afterwards the cells were washed 1 – 3 times with 1ml PBS/0.5% BSA and finally resuspended in 480µl PBS/0.5% BSA + 20µl of anti-rabbit IgG Microbead slurry (MACS). The mixture was incubated for 15min at 4°C under constant agitation, washed with PBS/0.5% BSA (300xg for 10min), resuspended in 500µl of the same buffer and applied to a column fixed to a magnet. Before the cells were applied to the column, it was equilibrated with 2ml of PBS/0.5% BSA. Bound cells were washed with 3ml PBS/0.5% BSA. Furthermore, 1ml PBS/0.5% BSA was applied to the bound cells, the column was removed from the magnet and the cells were collected in a tube. Eluted cells were washed with IPL-41/3% FCS + 50µg/ml Penicillin/Streptomycin and mixed with adherent *Sf*9 cells. 5 days "post infection" these cells were stained and analyzed by FACS.

3.10.8 *Fluorescence microscopy*

Adhesion slides (Marienfeld, 7.6 x 1 x 2.6mm) were washed with water. 20µl of insect cell suspension (2 – 3 days post infection) was spottet onto the slide at the center of drawn circles and incubated for 2h at 27°C for allowing cell adherence. Following procedures were performed at room temperature. Before and after cell fixation with 20µl 4% paraformaldehyde, the cells were carefully washed with PBS. Then 20µl of PBS/1% Triton was added to the cells and incubated for 10min, followed by a washing step with PBS. For reducing background signals, the cells were incubated for 30min with 30µl blocking solution (PBS/20% FCS). Anti-HRP antibody was diluted 1:500 and incubated with the cells for 45min, followed by three washing steps with PBS. Anti-rabbit IgG conjugated with PE was diluted 1:500 and incubated with the cells for 20min. Finally the cells were washed three times with PBS.

3.11 Glycosyltransferases

3.11.1 *ß1,2-XylT*

3.11.1.1 Positive control: Recombinant expression of plant ß1,2-XylT in insect cells

3.11.1.1.1 Determination of an optimal host strain for ß1,2-XylT expression

Different insect cell lines (*Sf*9, *Sf*21, Mimic, BTI-*Tnao*38 and Hi5) were grown for three days under standard conditions (chapter 3.2.3), harvested and diluted with water to a concentration of $3.4x10^6$ cells/ml. 0.5 to 2µl of each cell suspension were spottet onto nitrocellulose membrane (chapter 3.1.5). In addition, SDS-PAGE and Western blot was done in parallel by applying $3.4x10^4$ cells per lane. α1,3-fucosylated N-glycans were detected by the incubation with anti-HRP antibody (1:2000) followed by alkaline phosphatase conjugated anti-mouse IgG (1:2000).

3.11.1.1.2 Cloning and recombinant expression of ß1,2-XylT from A. thaliana

The plant ß1,2-XylT was amplified by PCR using a plasmid as template (gift from Doz. DI Dr. R. Strasser) and the primer set 5'AAGCAGTGGTATCAAATGAGTAAACGGAATCCGAAG3' (forward), 5'AAGCAGTGGTATCAATTAGCAGCCAAGGCTCTTC3' (reverse). Recombination sites are underlined. After DNA purification (chapter 3.4.2) the 1605bp gene was precipitated with EtOH (chapter 3.4.5) and recombined with pENTR1A vector containing the flanking sequences 5'TTGATACCACTGCTT3' inserted by an inverse PCR. The inverse PCR was performed using isolated pENTR1A vector as template and the primer set 5'TTGATACCACTGCTTATCTAGACCCAGCTTTCTT3', 5'TTGATACCACTGCTTGGTTCCTTTAAAGCCTGCT3'.

The recombination reaction contained In-Fusion reaction buffer, 100ng amplified pENTR1A, ~250ng amplified ß1,2-XylT gene (~4.5 molar excess) and In Fusion enzyme in a total volume of 10µl. The reaction was incubated for 15min at 37°C, followed by 15min at 50°C. After cooling down on ice, 5µl was transformed

chemically (chapter 3.5.2) into NEB10ß *E. coli* cells. Isolated plasmids were sequenced and further used for the recombination with linearized baculovirus genome *in vitro* as described in chapter 3.8.1. The recombinant expression in insect cells is described in chapter 3.7.3.

3.11.1.1.3 Characterization of ß1,2-XylT (A. thaliana) expressing insect cells

Western Blot: 6 days post infection *Sf*9 cells were harvested, washed three times with PBS buffer (1000xg for 5min) and finally resuspended in 100µl PBS. SDS-PAGE and Western Blot was performed (chapter 3.1.4). For testing the modified glycans on virus proteins, such as gp64, infected cells were centrifuged at 3500xg for 7min and the supernatant was applied to SDS-PAGE/Western Blot. For immunodetection anti-HRP antibody and anti-mouse IgG conjugated to AP were diluted 1:2000.

HPLC: Monosaccharides from GFP and ß1,2-XylT expressing *Sf*9 cells were analyzed by HPLC. $1x10^7$ cells were washed three times with PBS, resuspended in 500µl water and homogenized using the UltraTurrax (3x at max. speed for 15sec). The sample was precipitated with four volumes of cold MeOH at -80°C for 30min. Monosaccharides from cell pellets (14500rpm for 30min at 4°C) were released and labelled with anthranilic acid (chapter 3.10.3).

In parallel, 100µl supernatant of GFP and ß1,2-XylT expressing *Sf*9 cells was precipitated 2x times with 400µl MeOH. Monosaccharides from protein pellets (14500rpm for 30min at 4°C) were released and labelled with anthranilic acid as described in chapter 3.10.3.

FACS: Single viruses containing the ß1,2-XylT (*A.c.* XylT1 – 3) was isolated by performing a plaque assay (chapter 3.8.3) and used for re-infection of *Sf*9 cells. 3 days after post infection these cells were labelled for FACS analysis using anti-HRP antibody and anti-rabbit IgG conjugated to PE (chapter 3.10.6).

Fl microscopy: Sample preparation for Fluorescence microscopy was performed as written in chapter 3.10.8.

3.11.1.1.4 ß1,2-XylT activity assay (radioactivity)

The enzyme activity of the recombinant ß1,2-XylT from *A. thaliana* (gift from Univ.-Prof. Dr. L. Mach) was determined in 20µl reaction mixture containing 100mM MES (2-(N-morpholino)ethanesulfonic acid), pH 7.0, 10mM $MnCl_2$, 0.1% Triton X-100, 5nmol UDP-[^{3}H]-Xyl (5000cpm/nmol; Sigma-Aldrich), ~10nmol acceptor glycan (MM-PA, TAKA-GP, GnGn-OS, GnGnF^6, GnGn-PA; Figure 24), 4ng enzyme solution and 0.6µg BSA at 37°C for 4h. In addition, isolated microsomes from snail tissue (chapter 3.1.7) were used as enzyme source. The enzyme activity assay was further applied to an ion exchange chromatography (Dowex AG 1x8) for removing excess substrate, BSA and the enzyme. The flow through contained the (modified) acceptor substrate and released [^{3}H]-Xyl, so a negative control was done without the acceptor.

3.11.1.1.5 ß1,2-XylT activity assay (HPLC)

The enzyme activity of the recombinant ß1,2-XylT from *A. thaliana* (gift from Univ.-Prof. Dr. L. Mach) was performed under same conditions as described in chapter 3.11.1.1.4, but with the double amount of recombinant enzyme (8ng), ~10nmol GnGn-PA and 8nmol unlabelled UDP-Xyl as donor substrate overnight at 37°C. In addition, isolated microsomes from snail tissue (chapter 3.1.7) were used as enzyme source. The enzyme activity assay was measured by HPLC (chapter 3.10.2).

3.11.1.2 Screening for a snail ß1,2-XylT

3.11.1.2.1 FACS (Cell sorting)

Labelling of infected insect cells for FACS analysis was done as described in chapter 3.10.6. *Sf*9 cells were infected with recombinant baculovirus containing the cDNA library of *A. lusitanicus*. 3 days post infection 2000 cells with the highest fluorescence signal were sorted in 500µl IPL-41/3% FCS + 50µg/ml Penicillin/Streptomycin. These cells were spread on 7.5x10^5 *Sf*9 adherent cells. After 8 days at 27°C the supernatant (7min at 3500xg) was used for a new infection of *Sf*9 cells. 3 days post infection, cells were labelled again and 2000 cells with the highest

signal intensity were used for a new cycle of infection. Cell sorting was repeated three times.

3.11.1.2.2 Homology-based PCR

Highly conserved regions were identified by alignment of plant ß1,2-XylT genes from *Arabidopsis thaliana* (AJ272121.1), *Hordeum vulgare* (AM179853.1), *Populus alba* (AJ891042.1) and *Solanum tuberosum* (AM179855.1). Based on conserved nucleotide sequences, sense and antisense degenerated primers (5'GATTYGAGTAYGCAAAYCTCTTYCA3' and 5'GTGWGTBAGWCCWGCB CCATGAGC3') were designed. In parallel, based on the highly conserved amino acid sequences (KP/AWPI/R and DWYSAY) degenerated primers (5'AARCCNTGGCCNATH3' and 5'RTANGCNSWRTACCARTC3') were designed and used for screenings. Standard homology-based PCR was performed in 30µl solution containing 10ng of double-stranded cDNA synthesized from *A. lusitanicus*, *A. fulica*, *B. glabrata* or Bge cells, 3.8µM degenerated primers, 1.5mM $MgCl_2$, 200µM dNTP, and 1.5U of HybriPol DNA Polymerase (Bioline). The PCR program started with an initial denaturation step at 94°C for 4min, followed by 35 cycles (94°C 30sec, 55°C 40sec, 72°C 1min) and a final incubation at 72°C for 5min. In several PCRs the Mg^{++} concentration was changed ranging from 0.5mM to 2mM, as well as the annealing temperature was tested between 50°C to 60°C. Additionally, cDNA synthesized from total RNA, which was isolated from different snail tissues, was tested.

3.11.1.2.3 Dot blot of isolated microsomes from A. lusitanicus and rabbit brain

1.5 to 3µg microsome protein were applied onto a nitrocellulose membrane (chapter 3.1.5). As negative control 2µg fetuin and as positive control 2µg bromelain were spottet, respectively. Anti HRP antibody and anti-mouse IgG conjugated to AP were diluted 1:1000.

3.11.2 *ppGalNAcT*

3.11.2.1 Homology-based PCR

*By alignment of ppGalNAcT genes from Homo sapiens 2 (NP_004472), Mus musculus (NP_038842), Caenorhabditis elegans GLY-3 (NP_498722) and Drosophila melanogaster (AAQ56699) highly conserved regions were identified. Based on the conserved amino acid sequences WGGEN and VWMDEY/F sense and antisense degenerated primers (5'TGGGGWGGWGARAACYT3' and 5'TGWAYTCRTCCATCCADAC3') were designed. Homology-based PCR was performed in 30µl solution containing 10ng of double-stranded cDNA synthesized from Bge cells, 2.6µM degenerated primers, 1.5mM $MgCl_2$, 200µM dNTP, and 1.5U of HybriPol DNA Polymerase (Bioline). The PCR program started with an initial denaturation at 94°C for 4min, followed by 35 cycles (94°C 30sec, 55°C 40sec, 72°C 1min) and a final incubation at 72°C for 5min. The resulting 187 bp product was purified by NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel), ligated with the pGEM-T Easy vector (Promega) and sequenced to examine identity of the fragment.**

3.11.2.2 Isolation of the full-length ppGalNAcT gene

The complete sequence of this novel ppGalNAcT gene was isolated by inverse PCR of self-ligated cDNA, as well as a 5' and 3' RACE PCR, respectively.
Inverse PCR of self-ligated cDNA: Double-stranded cDNA (3.5µg) was precipitated, resuspended in NEBuffer 4 and incubated with DNA Polymerase I, Large (Klenow) Fragment in a total volume of 25µl for 15min at 25°C. After DNA purification and EtOH precipitation the cDNA was treated with 10U of T4 Polynucleotide Kinase in a total volume of 30µl for 30min at 37°C. The enzyme was inactivated at 65°C for 20min. T4 DNA Ligase was used for the self-ligation reaction (16°C, overnight). Correct self-ligated cDNAs coding for the ppGalNAcT gene were amplified by inverse PCR using the primer set 5'GAAACACAAAGCGTGCAGCAGAAG3' and 5'CACTGCCAAACACGGAAGGATATC3'. Following an initial denaturation step at 94°C for 4min, 35 x PCR cycles (94°C 30sec, 60°C 40sec, 72°C 1.5min) resulted in

the amplification of about 2000 bp fragments. These amplified products were purified and ligated with the pGEM-T Easy vector for sequencing.

*RACE-PCR: 5' and 3' Ready RACE cDNA was synthesized from 1µg of total RNA by using the SMARTer™ RACE cDNA Amplification Kit (Takara). The complete 5' end of the novel ppGalNAcT gene was amplified by using the Phusion High-Fidelity DNA Polymerase (Thermo Scientific), 5' Ready RACE cDNA as template, the gene specific primer 5'CATCCAGACTTCTGCTGCACGCTTTGTG3' and the 10X Universal Primer A Mix provided by the kit. After 6x PCR cycles of a 2-step PCR (98°C 10sec, 72°C 1min 20sec), 30x PCR cycles of a 3-step PCR (98°C 10sec, 69°C 30sec and 72°C 1min 20sec) followed. The amplified 5' end was purified and ligated with pUC19 vector for sequencing. A nested PCR resulted in the amplification of the complete 3' end using ExTaq Polymerase (Takara) and 3' Ready RACE cDNA as template. After an initial denaturation step at 94°C for 3min, 20x PCR cycles (94°C 30sec, 69°C 30sec, 72°C 1min 20sec) were performed with the gene specific primer 5'CACAAAGCGTGCAGCAGAAGTCTGGATG3' and the Universal Primer A Mix. Then 40x PCR cycles (94°C 30sec, 68°C 30sec, 72°C 1min) followed with the nested primer 5'GGGAATGACTGTATGGACACTATGGGCC3' and the Nested Universal Primer A provided by the kit. The amplified 3' end was purified and ligated with pGEM-T Easy vector for sequencing.**

3.11.2.3 Recombinant expression, purification and Western Blot analysis

In order to express soluble ppGalNAcT, a cDNA fragment without the putative cytoplasmic tail and the transmembrane domain (ΔTM ppGalNAcT) was amplified by using the primer set 5'GGG<u>GAGCTC</u>GGAGATGATCAAAGTGAGTTTG3' (SacI restriction site underlined) and 5'GGG<u>GGTACC</u>CTATCTGTTTTTACTTAAA GAAAATGTCCAC3' (KpnI restriction site underlined). The purified ΔTM ppGalNAcT gene was ligated with the pGEM-T Easy vector. Resulting plasmids were confirmed by sequence analysis. After SacI/KpnI digestion the ΔTM ppGalNAcT gene was purified and ligated to the transfer vector pVT-Bac His1. To verify the gene sequence and the correct insertion into the cloning site, this construct was sequenced. 2µg of the recombinant construct (containing the honeybee melittin secretion signal, a 6 x His tag and the truncated ppGalNAcT gene) were co-transfected with 300ng linearized BaculoGold DNA (BD Biosciences) into Sf9 insect cells to produce

*recombinant baculovirus (chapter 3.8.1). Infection and recombinant expression of the novel ppGalNAcT was performed as described in chapter 3.7.3. The soluble enzyme in the culture medium was purified on a Ni-NTA agarose (Qiagen) column using a gradient from 10 to 250mM imidazole in buffer (25mM Tris/HCl pH 7.2, 300mM NaCl) at a flow rate of 0.5ml/min in 30min. Eluted recombinant enzyme was detected at 280nm. Aliquots of protein containing samples were applied to SDS-PAGE and detected either with Coomassie staining or immunoblotting on nitrocellulose membrane. Immunodetection was performed by incubation with Penta-His monoclonal antibody (Qiagen; dilution 1:2000) followed by alkaline phosphatase conjugated anti-mouse IgG from goat (Sigma-Aldrich; dilution 1:3000).**

3.11.2.4 Characterization of the Bge ppGalNAcT

*Enzyme activity of the recombinant ppGalNAcT was determined in 20µl reaction mixture containing 50mM MES (2-(N-morpholino) ethanesulfonic acid), pH 7.0, 10mM $MnCl_2$, 0.1% Triton X-100, 40nmol UDP-GalNAc (Sigma-Aldrich), 20µg acceptor peptide (Cellmano Biotech Co) and 2µl enzyme solution (protein concentration 5.75µg/ml) at 37°C for 45min. The peptides were chosen randomly: Muc1a (APPAHGVTSAPDTRPAPGC) and 1a' (AHGVTSAPDTR) were taken from (Wandall et al, 1997), Muc2 (PTTTPITTTTTVTPTPTPTGTQTK) from (Wang et al, 2003) and Muc5Ac (GTTPSPVPTTSTTSAP) from (Ten Hagen et al, 1999). The reaction was terminated by boiling at 96°C for 5min and analyzed by HPLC on a reversed phase C18 column (4.6×250 mm, 5µm,Thermo Scientific) in 0.1% trifluoroacetic acid in water, applying a linear gradient from 10 to 30% of eluent (0.1% trifluoroacetic acid in acetonitrile) in 25min at a flow rate of 1ml/min. Detection was carried out at 220nm. The long term incubation was carried out for 24h analogous to the standard assay with additions of 10nmol UDP-GalNAc and 2µl of enzyme solution every 3 hours. For the analysis of the biochemical parameters the standard assay conditions using Muc2 as acceptor peptide were modified as follows:** For determination of the manganese optimum the concentration of $MnCl_2$ was varied from 0–300mM, for the determination of cation requirement the standard assay was carried out without any cation addition or in the presence of 10mM of EDTA, Mn^{2+}, Mg^{2+}, Ca^{2+}, Co^{2+}, Cu^{2+}, Ni^{2+}, or Ba^{2+}.

Enzyme stability was tested by storing the enzyme under different conditions for two days. On the one hand, 5/10% MeOH, 5/10% AcCN and 5% glycerine was added and on the other hand the enzyme was stored at -20°C, 4°C, 25°C, 37°C and 50°C. In addition, the enzyme was lyophilized. The enzyme activity assays with the treated ppGalNAcT solutions were carried out under standard conditions.

For the determination of the pH optimum several buffer systems were used under standard conditions ranging from pH 3.5 to 9.0 (acetate/NaOH: 3.5 – 5.0; citrate/NaOH: 4.5 – 6.0; MES: 5.5 – 7.0; phosphate: 6.0 – 7.5; Tris/HCl: 7.0 – 9.0) in pH 0.5 steps.

Donor substrate specificity was tested using 40nmol of UDP-GalNAc, UDP-Gal, GDP-Fuc, UDP-Xyl, UDP-GlcNAc and CMP-NANA, respectively, under standard conditions. In addition, it was tested if the sugars were transferred to monoglycosylated Muc2 peptide, which had been isolated by HPLC.

Kinetic data were acquired using 1:4 diluted enzyme solution and Muc2 at eleven different concentrations ranging from 0.01 to 0.82mM, with UDP-GalNAc kept constant at a concentrationof 2mM. Similarly UDP-GalNAc was varied between concentrations from 0.01 to 2mM while Muc2 was kept constant at 0.82mM. K_M*-values were obtained from Lineweaver-Burk plots. All quantitative values were calculated from the area of HPLC patterns. Each assay was carried out at least in duplicate with appropriate controls.*

*For determination of O-glycosylation 34.5ng of recombinant ppGalNAcT were treated with 1mU of O-glycosidase (Sigma-Aldrich) in a total volume of 20µl in the presence of 50mM sodium phosphate pH 5.0 for 3h at 37°C and applied to SDS-PAGE followed by lectin blot analysis. Lectin blot was done by incubation with biotinylated peanut agglutinin (Vector Laboratories; dilution 1:1333) and detected by streptavidin-alkaline-phosphatase conjugate (Vector Laboratories; dilution 1:1333).**

The acceptor substrate specificity was determined using 22 different peptides (Table 8; Table 9) under standard conditions. On the one hand, an optimal peptide core (-3 to +3 relative to the O-glycosylation site) for human ppGalNAcT-2 (Gerken et al, 2006) was designed, which is designated as CHT1 (APPAHPGPTPGPRPAPG). Peptides (CHT2 – CHT15) contained amino acid substitutions of this core region, so the product conversion of these peptides were correlated to the original CHT1 product amount. CHT1 product conversion was assumed as 100%.

On the other hand, a peptide core described as being optimal for human ppGalNAcT-1 (Gerken et al, 2006) was synthesized and named as CHT16 (APPAHDFVTPAPRPAPG). CHT17 – CHT22 are substituted with amino acids, preferred from human ppGalNAcT-2, and product conversion was calculated in relation to CHT16.

---

* Methods were described in Taus et al, 2013

3.11.3 *T-Synthase*

3.11.3.1 Homology-based PCR

Highly conserved regions were identified by alignment of T-Synthases from *H. sapiens* (AAF81981.1), *M. musculus* (AAF81982.1), *C. elegans* (AAG36940.1) and *D. melanogaster* (AAF52724.1). Based on the conserved amino acid sequences WA/GK/QRCNKV/L/YV/I/L and WFL/MKADDD sense and antisense degenerated primers (5'ACBTGGGSCMAGCGKTGYAAYAA3' and 5'GTRTCRTCRTCDGCY TTCADRAAC3') were designed. Homology-based PCR was performed in 25µl solution containing 10ng of double-stranded cDNA synthesized from *A. lusitanicus, A. fulica, B. glabrata* or Bge cells, 2.0µM degenerated primers, 200µM dNTP, and ExTaq Polymerase (Takara). The PCR program started with an initial denaturation at 94°C for 3min, followed by 40 cycles (94°C 30sec, 60°C 40sec, 72°C 1min) and a final incubation at 72°C for 7min. The resulting product was purified, ligated with the pGEM-T Easy vector (Promega) and sequenced to examine identity of the fragment.

3.11.3.2 5' RACE PCR

The complete 5' end of the novel T-synthase gene was amplified by performing a nested PCR. The first PCR was done by using Phusion High-Fidelity DNA Polymerase (Thermo Scientific), 5' Ready RACE cDNA as template, the gene specific primer 5'CATGAACCAATCAGCATCATTGAAGTGATGCTC3' and the 10X Universal Primer A Mix provided by the kit. The PCR program started with an initial denaturation step at 98°C for 30sec, followed by 6x cycles (98°C 10sec, 72°C 1min 20sec) and further 30x cycles (98°C 10sec, 69°C 30sec, 72°C 1min 20sec). The nested PCR was performed with 1µl template of the PCR reaction described above, Phusion High-Fidelity DNA Polymerase (Thermo Scientific), the gene specific primer 5'CGTACCGGAAGCCGCGCATGGTTTTG3' and the nested Universal Primer A provided by the kit. 40x PCR cycles (98°C 10sec, 64°C 40sec, 72°C 40sec) followed an initial denaturation step of 98°C for 30sec. The amplified fragment was

ligated with p-GEMT vector, transformed chemically into NEB10ß *E. coli* cells and sequenced after plasmid isolation was performed.

3.11.3.3 3' RACE PCR

3' Ready RACE cDNA was synthesized using SMARTer™ RACE cDNA Amplification Kit (Takara). 3' RACE PCRs were performed with different DNA Polymerases, such as ExTaq Polymerase (Takara), Phusion High-Fidelity DNA Polymerase (Thermo Scientific) and HybriPol DNA Polymerase (Bioline), various annealing temperatures ranging from 58°C to 68°C and 6 gene specific forward primers (TS1 – TS6; Figure 10; Figure 11). As reverse primers the Universal Primer A Mix, as well as the Nested Universal Primer A, both provided by the kit, were used. Fragments amplified in PCRs using primers located at the extreme ends, such as with TS1 and Universal Primer A Mix, represented the template for nested PCRs using TS2 – TS6 and the Nested Universal Primer A in every combination. Up to 3 subsequent nested PCRs were done.

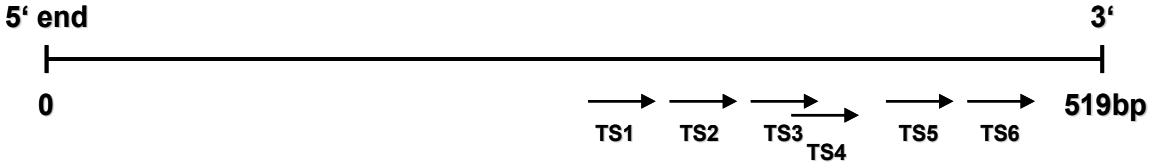

Figure 10: Scheme of the complete 5' end (519bp) of the T-synthase of *B. glabrata*. Arrows (TS1 – TS6) represent the primers used for 3' RACE (nested) PCRs

ATGAACGCACCCCCAAAACAGTTTCTACTCGGCCTGCTGGCCGGATTGTCCTTGA
CTTTTCTGATCTACAGCTATGTCCACCTCAGCTACCGAGAGCCACCACACCCGAC
AGAATTCCGGTCAACTAGCCTTCTAGACAGCTTGTCAAAAAAGCAGGACAGCCAC
GAGCAATTTGATCAGGTCATCGCACGTTCCGCTCTTCCGGAAGTCCATCTTGATG
AGGATAAGCACATGCATCATGACG**ATGACAAGGAGGCCAAACAACTTGCTG**AG
AAAATTAAAGTTCTGGTCTGGGTCATGACCAACCCAAACAATC**TGGAGAAGAAG**
GCCAAGATCGTAAAGGAGACTTGGGGCAAGCGCTGTAATAA**GGTGATCTTTTT**
CAGCTCAGAGACCAACAATGACTTTCCAACTATTGGATTGGA**CGTACCTGAA**
GGACGCGAGCACCTGACAGCCAAAACCAT**GCGCGGCTTCCGGTACGTGTA**CG
AGCACCACTTCAATGATGCTGATTGGTTCATGAAG¶

Figure 11: Sequence of the complete 5' end (519bp) of the T-synthase from *B. glabrata*. Underlined sequences represent the primers (TS1 – TS6) used for 3' RACE (nested) PCR

3.12 C-type lectin from *A. lusitanicus*

3.12.1.1 Cloning and recombinant expression in insect cells

The full-length gene of a C-type lectin was discovered by sequencing 30 *E. coli* colonies containing the pBacPAK8-cDNA library from *A. lusitanicus*. The plasmid of the correct clone was isolated and amplified by inverse PCR using the primer set 5`CCCACTAGTTAATCTTCCATGACAGCCGAG3' (forward), 5`CCCACTAGTATGATGATGATGATGATGCTTAACGTCAGCCTTTCTC3' (reverse) and 10ng recombinant plasmid as template. SpeI restriction sites are underlined. By this amplification a His-tag was introduced to the 3' end of the C-type lectin gene. After amplification, the linearized vector was purified from agarose gel, EtOH precipitated and restricted with SpeI for 3h at 37°C. SpeI digested vector was purified and used directly for self-ligation reaction. Following transformation, plasmid isolation and sequencing, the recombinant plasmid was co-transfected with linearized BaculoGold DNA (BD Biosciences) to produce recombinant virus (chapter 3.8.1).

To introduce a His tag to the 5' end of the C-type lectin gene, it was ligated in parallel with the transfer vector pVT-Bac-His1. For this aim, the C-type lectin gene was amplified by PCR using the primer sets 5' CCCGCTAGCATGATTCTTGCCGTCTTGG3' and 5'CCCGAATTCTTACTTAA CGTCAGCCTTTC3' (full-length) or 5'CCCGCTAGCCAATACCAGGAAAA CTGC3' and 5'CCCGAATTCTTACTTAACGTCAGCCTTTC3' (ΔTM), respectively. NheI and EcoRI restriction sites are underlined. Amplified genes were purified from agarose gel, digested with NheI and EcoRI for 2.5h at 37°C, purified and ligated with pVT-Bac-His1 vector, which had been digested with NheI and EcoRI, too. Before the ligation reaction was performed at 14°C overnight, the vector and the inserts were precipitated with EtOH. Following chemical transformation into NEB10ß *E. coli* cells and subsequent sequencing, isolated recombinant plasmids containing the full-length or ΔTM C-type lectin gene were co-transfected with BaculoGold DNA (BD Biosciences) (chapter 3.8.1). Recombinant expression in insect cells was done as described in chapter 3.7.3.

3.12.1.2 Cloning and recombinant expression in *P. pastoris*

Isolated pPICZα was digested with EcoRI and purified from agarose gel. The vector was treated with Antarctic Phosphatase (15min at 37°C followed by 10min at 65°C), purified and EtOH precipitated. The C-type lectin was amplified by PCR with and without the transmembrane domain by using recombinant plasmid as template and the primer sets 5'CCCGAATTCATGATTCTTGCCGTCTTGG3' and 5'CCCGAATTCTTACTTAACGTCAGCCTTTC3' (full length gene) or 5' CCCGAATTCCAATACCAGGAAAACTGC3' and 5'CCCGAATTCTTACTTAA CGTCAGCCTTTC3' (ΔTM), respectively. EcoRI restriction sites are underlined. The amplified genes were purified from agarose gel, EtOH precipitated and EcoRI digested for 1.5h at 37°C. After removing salts and restriction enzyme by column purification, EcoRI digested pPICZα and C-type lectin gene (full-length/ ΔTM) were ligated at 14°C overnight. Following chemically transformation into NEB10ß *E. coli* cells and sequencing, isolated plasmids were linearized by using PmeI restriction enzyme. The recombinant DNA was purified and transformed into electrocompetent yeast cells, as described in chapter 3.6.2. The recombinant expression in yeast cells is described in chapter 3.7.2.

3.12.1.3 Cloning and recombinant expression in *E. coli*

pET26b vector was isolated from an *E. coli* ONC, digested with EcoRI and SacI, purified and precipitated with EtOH. The C-type lectin gene was amplified by PCR with and without the transmembrane domain using recombinant plasmid as template and the primer set 5'CCCGAATTCGATGATTCTTGCCGTCTTGG3' and 5'CCCGAGCTCTTAACGTCAGCCTTTCTC3' (full-length) or 5'CCCGAATTCG CAATACCAGGAAAACTGC3' and 5'CCCGAGCTCTTAACGTCAGCCTTTCT C3' (ΔTM). EcoRI and SacI restriction sites are underlined. The amplified genes were purified from agarose gel, EtOH precipitated and EcoRI/SacI digested for 2h at 37°C. After DNA purification the C-type lectin genes were ligated with the digested pET26b vector at 14°C overnight. As an intermediate step, the plasmids were first transformed into NEB10ß *E. coli* cells and sequenced after plasmid isolation. Controlled plasmids were further transformed chemically into BL21 *E. coli* cells and used for the recombinant expression. The procedure is described in in chapter 3.7.1.

3.12.1.4 Purification in the presence of detergents

Before the purification started, several detergents were tried to solubilize the recombinant C-type lectin. Pellets of infected BTI-*Tnao*38 cells (600xg for 10min) were resuspended in 30µl of different buffers (Table 3), vortexed and incubated for 20min at room temperature. Furthermore, the cell suspension was centrifuged at 6000rpm for 3min and 7µl supernatant were applied to SDS-PAGE/Western Blot. Anti-His antibody, as well as anti mouse IgG-AP were diluted 1:2000.

Table 3: Different buffers to solubilize the recombinant snail C-type lectin

1	2	3	4
50mMTris/HCl pH 9.0 + 1% Triton X-100	50mMTris/HCl pH 9.0 + 1% Tween 20	50mMTris/HCl pH 9.0 + 1% Triton X-100 + 0.5mM DTT	50mMTris/HCl pH 9.0 + 1% Triton X-100 + 0.1% SDS

Infected BTI-*Tnao*38 cells were washed 2x with PBS (1500xg for 10min) to remove the culture media completely (15mM His), resuspended in 4ml Lysis buffer (Table 4) and incubated for 30min at RT. 500µl Ni-NTA slurry (Qiagen) were filled in the column, washed with 3ml PBS and further equilibrated with 3ml Lysis buffer. Lysed cells were centrifuged (5000xg for 10min) and the supernatant was applied to the column. Flow through, Wash 1 (+1ml Lysis buffer; Table 4), Wash 2 (+1ml Wash buffer; Table 4) and Elution fractions (+2ml Elution buffer; Table 4) were collected and applied to SDS-PAGE/Western Blot.
This protocol was modified by incubating the sample with Ni-NTA for several hours at 4°C under constant agitation before washing and elution steps were performed, the reduction of imidazole concentration to 1mM in Lysis buffer or the substitution of Triton-X100 by IGEPAL CA-630.

Table 4: Triton X-100 containing buffers used for C-type lectin purification

Lysis buffer	Wash buffer	Elution buffer
50mMTris/HCl pH 9.0 300mM NaCl 1% Triton X-100 10mM imidazole	50mMTris/HCl pH 9.0 300mM NaCl 1% Triton X-100 20mM imidazole	50mMTris/HCl pH 9.0 300mM NaCl 1% Triton X-100 250mM imidazole

3.12.1.5 Purification without detergents

Several conditions were tested for solubilizing the recombinant C-type lectin without the presence of detergents. Pellets of infected BTI-*Tnao*38 cells (600xg for 10min) were resuspended in 500µl water, mixed with 300µl glass beads and vortexed roughly (10x for 20sec). In between the vortex steps, the sample was cooled on ice for 30sec. Furthermore, the cell lysate was sonicated (Duty Cycle = 30; Output Control = 3; 3min) and centrifuged at 500xg, 1000xg, 2000xg, 8000xg, 14100xg and 30000xg, respectively, for 5min. The supernatants of all samples were applied to SDS-PAGE/Western Blot. Anti-His antibody, as well as anti mouse IgG-AP were diluted 1:2000.

Supernatants containing the C-type lectin were used for protein purification. Lysed cells were adjusted to 50mM Tris/HCl, 300mM NaCl pH 7.1. The sample was applied to the column filled with Ni-NTA, equilibrated in Basic buffer (500µl; Table 5). Flow through, Wash 1 (+2ml Wash buffer 1; Table 5), Wash 2 (+2ml Wash buffer 2; Table 5) and Elution fractions (+2ml Elution buffer; Table 5) were collected and applied to SDS-PAGE/Western Blot. This protocol was modified by using 10x more Ni-NTA gel. Concentration of elution fractions by using CentriPreps (3kDa cut-off) sometimes was necessary to visualize protein bands by Western blot.

Table 5: Buffers without detergents used for C-type lectin purification

Basic buffer	Wash buffer 1	Wash buffer 2	Elution buffer
50mM Tris/HCl pH 7.1 300mM NaCl	50mM Tris/HCl pH 7.1 300mM NaCl 10mM imidazole	50mM Tris/HCl pH 7.1 300mM NaCl 20mM imidazole	50mMTris/HCl pH 9.0 300mM NaCl 250mM imidazole

3.12.1.6 Purification under denaturing conditions

Pellets of infected BTI-*Tnao*38 cells (600xg for 10min) were resuspended in 500µl Buffer B (Table 6), mixed with 300µl glass beads and vortexed roughly (10x for 20sec). In between the vortex steps, the sample was cooled on ice for 30sec. The cell lysate was sonicated (Duty Cycle = 30; Output Control = 3; 3min) and centrifuged at

8000rpm for 10min. The supernatant was mixed with Ni-NTA equilibrated in Buffer B (Table 6) at 4°C for 1h under constant agitation. The mixture was applied to the column and flow through, Buffer C (+4ml Buffer C; Table 6), Buffer D (+3ml Buffer D; Table 6) and Buffer E (+2ml Buffer E; Table 6) fractions were collected and applied to SDS-PAGE/Western Blot. Anti-His antibody was diluted 1:2000, whereas anti mouse IgG-AP were diluted 1:5000.

Table 6: Denaturing buffers used for C-type lectin purification

Buffer B	Buffer C	Buffer D	Buffer E
100mM NaH_2PO_4 10mM Tris/HCl 8M urea pH 8.5	100mM NaH_2PO_4 10mM Tris/HCl 8M urea pH 6.5	100mM NaH_2PO_4 10mM Tris/HCl 8M urea pH 5.8	100mM NaH_2PO_4 10mM Tris/HCl 8M urea pH 4.5

3.12.1.7 Purification using Mannan agarose

3x25ml BL21 cell cultures expressing either the full-length or the ΔTM C-type lectin were combined after 4h IPTG induction, centrifuged (11000xg for 5min), resuspended in 3ml Loading buffer (10mM imidazole, 1.25M NaCl, 20mM $CaCl_2$ pH 7.8) and sonicated on ice (Duty Cycle = 60; Output Control = 3; 2min). Lysed cells were centrifuged at 14500rpm for 30min and 10µl supernatant was applied to SDS-PAGE/Western Blot for checking of the solubility. The soluble ΔTM C-type lectin was dialyzed (3kDa cut-off) against 250ml Loading buffer. 2ml Mannan agarose was washed with Loading buffer and incubated with the dialyzed sample for 30min. Flow through, Wash (+6ml Loading buffer) and Elution fractions were collected and applied to SDS-PAGE, followed by a blot with anti-His antibody and anti-mouse IgG-AP (diluted 1:1500).

4 RESULTS

In summary, two major strategies were pursued to isolate novel glycosyltransferases from snail origin. On the one hand, protein purification using chromatographic procedures from homogenized snail tissue was planned to isolate desired enzymes. Several molecular protein characteristics, such as size, hydrophobicity and charge, should be utilized to get rid of unwanted proteins. Prior to protein purification, an enzyme activity assay was tried to establish to determine enzyme containing fractions.

On the other hand, molecular biological tools were used for the identification of genes encoding desirable enzymes. In this case, two strategies were followed. For the isolation of a snail ß1,2-XylT an expression cDNA library from *A. lusitanicus* was screened for the enzymes's product, so for ß1,2-xylosylated N-glycans exposed on the cell surface. ß1,2-xylosylation, as a result of the recombinant enzyme, was shown by the characterization of *Sf*9 cells expressing the *A. thaliana* ß1,2-XylT (positive control). Furthermore, an alignment of desired glycosyltransferase genes (ß1,2-XylTs; ppGalNAcTs; T-synthases) from different plant/animal species was performed to design degenerated primers specific to highly conserved motifs. If the nucleotide sequence was not conserved throughout the animal kingdom, conserved amino acid sequences were used for primer design. Subsequent the amplification of the corresponding snail fragment from cDNA libraries by homology-based PCR, the complete cDNA ends were isolated by 3'/5' RACE PCRs, respectively.

4.1 Gene diversity of snail cDNA libraries

Based on the fact that only scattered sequencing data from gastropods were available when the project started, the first step to isolate novel glycosyltransferases was to establish cDNA libraries from *A. lusitanicus*, *A. fulica*, *B. glabrata* and Bge cells. The quality, in terms of gene diversity, was tested by gel electrophoresis of synthesized cDNA and control PCRs of cDNA libraries. As a representative, the total number of genes from the *A. lusitanicus* cDNA library was calculated by sequencing several *E. coli* colonies. 36 colonies containing the pSMART2IFD cDNA library (*A. lusitanicus*) were sequenced (in total the cDNA library contained $2x10^5$ colonies) and six novel genes could be identified. About 50% inserts were coding for the 16S rRNA gene from *A. lusitanicus* and 33% inserts encoded a conserved hypothetical protein from *Clostridium cellulovorans*. Apparently, the cDNA library of *A. lusitanicus* was contaminated with cDNA of this anaerobic, spore forming, gram-negative bacterium. 17% of the sequenced colonies contained individual genes, meaning that the whole cDNA library of *A. lusitanicus* was coding for about 30000 different snail genes.

Before the cDNA was recombined with various vectors (pSMART2IFD, pGAPZ B and pBacPAK8), the quality was checked by gel electrophoresis. The double-stranded cDNA synthesized from *A. lusitanicus*, *A. fulica*, *B. glabrata* and the Bge cell line showed a broad smear ranging from 500bp to 10000bp indicating a high gene diversity (Figure 12).

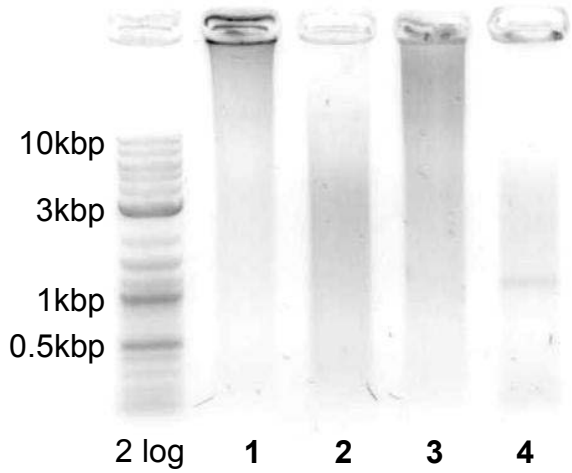

Figure 12: cDNA diversity of different snail species. (1) *A. lusitanicus* (2) *A. fulica* (3) *B. glabrata* (4) Bge cell line

As a further quality control the synthesized cDNA of all snail species was recombined with pSMART2IFD and transformed into NEB10ß *E. coli* cells. The insertion of individual genes was tested by colony PCR. 8 tested colonies of each cDNA library contained inserts of different size indicating useful gene diversity for screening procedures (Figure 13).

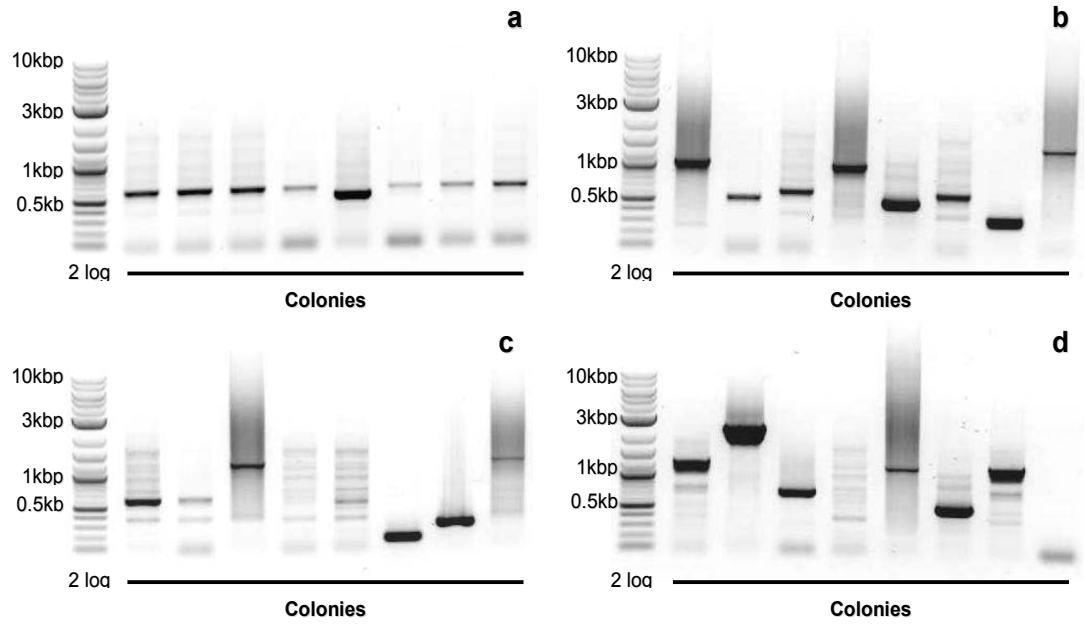

Figure 13: Colony PCR of 8 single *E. coli* colonies containing the pSMART2IFD cDNA library of different snail species. (a) *A. lusitanicus* (b) *A. fulica* (c) *B. glabrata* (d) Bge cell line

In addition to the cDNA libraries created in *E. coli* (pSMART2IFD), cDNA of *A. lusitanicus* was recombined with the yeast vector pGAPZ B, as well as with the transfer vector pBacPAK8 for the creation of recombinant baculovirus. The gene diversity of both cDNA libraries was tested by sequencing. Based on 24 sequenced colonies containing the pGAPZ B cDNA library, about 15% were coding for individual genes. In total the library comprised of $1.5x10^5$ yeast cell colonies. By the assumption that no gene diversity was lost during yeast cell transformation, the cDNA library contained about 23000 different genes.

About 33% *E. coli* colonies containing the pBacPAK8 cDNA library from *A. lusitanicus* were coding for individual genes. 67% of colonies coded for the 16S rRNA gene. This was estimated by sequencing 30 different colonies. In addition, the gene diversity within the recombinant baculovirus was tested by control PCR using

isolated baculovirus genome as the template. The broad smear ranging from 500bp to 9000bp (Figure 14), similar to the original pBAcPAK8 cDNA library, indicated that no significant loss of diversity due to the recombination procedure has been occurred. Cell sorting methods using FACS/MACS were tested with the *A. lusitanicus* cDNA library expressed in insect cells (pBacPAK8). The isolation of specific glycosyltransferases failed, so a further cDNA library of another snail species was not created with the pBacPAK8 vector. The cDNA library of *A. lusitanicus* created in yeast cells represent a capable sample for future protein screenings by containing about 23000 snail genes. For the isolation of novel glycosyltransferases, non-ligated cDNA and cDNA libraries created in pSMART2IFD were used as template for homology-based and RACE PCR. Table 7 summarizes the cDNA libraries of all snail species that had been created during this work.

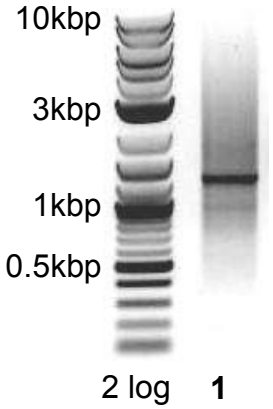

Figure 14: Control PCR of recombinant baculovirus containing the cDNA library of *A. lusitanicus*; (1) cDNA inserted in the virus genome

Table 7: cDNA libraries created in this work. * Created and stored in *E. coli* cells too

Snail species	***E. coli* cDNA library (pSMART2IFD)**	***P. pastoris* cDNA library (pGAPZ B*)**	**Baculovirus cDNA library (pBacPAK8*)**
A. lusitanicus	$2x10^5$ colonies 17% (30000) genes Sequencing; Control PCR	$1.5x10^5$ 15% (23000) genes Sequencing; Control PCR	$2.0x10^4$ 33% (7000) genes Sequencing; Control PCR
A. fulica	Control PCR	Not created	Not created
B. glabrata	Control PCR	Not created	Not created
Bge cell line	Control PCR	Not created	Not created

4.2 ß1,2-XylT

For the recombinant expression of the plant ß1,2-XylT (chapter 3.11.1.1), as well as for the functional screening of the snail ß1,2-XylT (chapter 3.11.1.2), insect cells were used as hosts. In comparison to yeast cells, capable acceptor glycans for the recombinant ß1,2-XylT, based on GlcNAcT-I activity, occur in insects. Thus, the cDNA library of *A. lusitanicus* created in yeast, was not screened for this enzyme. It represents a useful sample to screen for alternative snail proteins, because more than 23000 genes were conserved within this library.

4.2.1 *Positive control: Characterization of insect cells expressing the plant ß1,2-XylT*

In order to establish a reliable screening procedure for the desired snail ß1,2-XylT, as a positive control the ß1,2-XylT from *A. thaliana* was expressed recombinantly in insect cells. Using several analytical techniques, it was tested if N-glycans from endogeneous proteins were ß1,2-xylosylated, as soon as the plant ß1,2-XylT was expressed and active in these host cells.

ß1,2-xylosylated N-glycans were detected by anti-HRP antibody. Besides its specificity for ß1,2-Xyl linked to the ß-Man, anti-HRP additionally binds to core α1,3-Fuc of N-glycans (Kurosaka et al, 1991). Latter modification is typical for insects, so first the optimal insect cell line had to be chosen in terms of the least native core α1,3-fucosylation to avoid strong background signal. In summary, the cell lysate ($3.4x10^4$) of five different strains (*Sf*9, *Sf*21, Mimic, BTI-*Tnao*38 and Hi5) were separated by SDS-PAGE, blotted and labelled with anti-HRP antibody. *Sf*9 (Summers and Smith 1987), *Sf*21 (Vaughn et al, 1977) and Mimic cells were obtained from *Spodoptera frugiperda*. Mimic cells represent a transgenic cell line that constitutively express five mammalian glycosyltransferases resulting in the synthesis of biantennary, terminally sialylated N-glycans (Hollister et al, 2002). BTI-*Tnao*38 and Hi5 cells were isolated from *Ascalapha odorata* and *Trichoplusia ni*, respectively, and produce high amounts of recombinant protein in comparison to *Sf*9 cells . (Hashimoto et al, 2010; Wickham & Nemerow, 1993).

BTI-*Tnao*38 and Hi5 cells showed the strongest signal based on anti-HRP antibody labelling, indicating that core α1,3-Fuc was a frequent modification of these cell lines. *Sf*9, *Sf*21 and Mimic cells exhibited less amounts of α1,3-fucosylated N-glycans, interpreted from the low signal shown in the Western blot (Figure 15). The same result was achieved by Dot blots (data not shown). These results correlated with previous studies, where it has been shown that a considerable portion of N-glycans from BTI-*Tnao*38 and Hi5 cells carry both core α1,3- and α1,6-linked fucose, whereas *Sf*9 and Mimic cells mainly have α1,6-linked fucose (Palmberger et al, 2011; Palmberger et al, 2012).

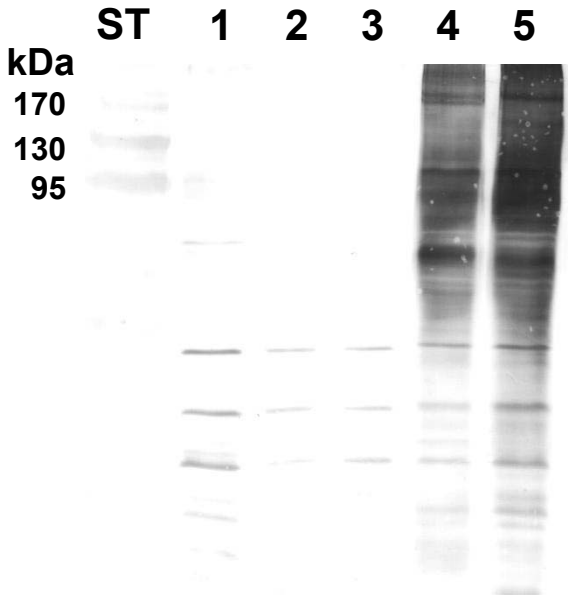

Figure 15: Western Blot using anti-HRP antibody of different insect cell lysates ($3.4x10^4$ cells per lane). (1) *Sf*9; (2) *Sf*21; (3) Mimic; (4) BTI-*Tnao*38; (5) Hi5

Based on less core α1,3-fucosylation detected with anti-HRP antibody, recombinant expression of the plant ß1,2-XylT and subsequent characterization of this positive control was performed using *Sf*9 cells. Recombinant DNA of 3 single viruses isolated by plaque assay were tested for correct insertion of the ß1,2-XylT gene within the baculovirus genome. By control PCR it could be confirmed that all 3 single viruses contained the gene from *A. thaliana* (1605bp) (Figure 16). Recombinant baculoviruses containing the plant ß1,2-XylT were designated *A.c.* XylT1 - 3.

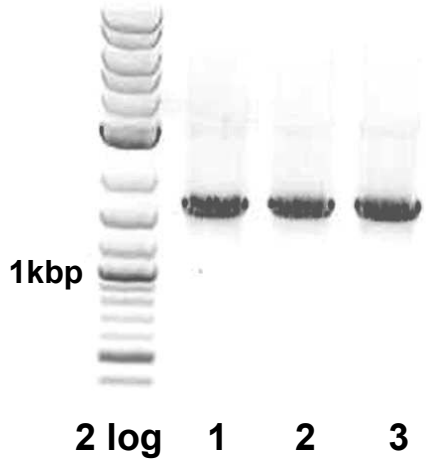

Figure 16: Control PCR of recombinant baculoviruses containing the ß1,2-XylT gene from *A. thaliana*. (1), (2), (3) Isolated virus DNA from 3 different plaques (*A.c.* XylT1 – 3)

Several analytical tools (FACS, Western blot, HPLC, Fluorescence microscopy) were used to determine N-glycan ß1,2-xylosylation of *Sf*9 cells infected by *A.c.* XylT1 - 3.

In comparison to the negative control (*A.c.* NPV), *Sf*9 cells expressing a nucleoprotein from influenza virus, *A.c.* XylT infected cells were bound stronger by anti-HRP antibody conjugated to PE (Figure 17). Infection with *A.c.* XylT1 (Figure 17; red line) yielded in the highest signal, indicating the strongest ß1,2-XylT expression and ß1,2-xylosylation of native N-glycans. For subsequent characterization of the positive control using Western blot, HPLC and fluorescence microscopy, *A.c.* XylT1 was used for *Sf*9 infection.

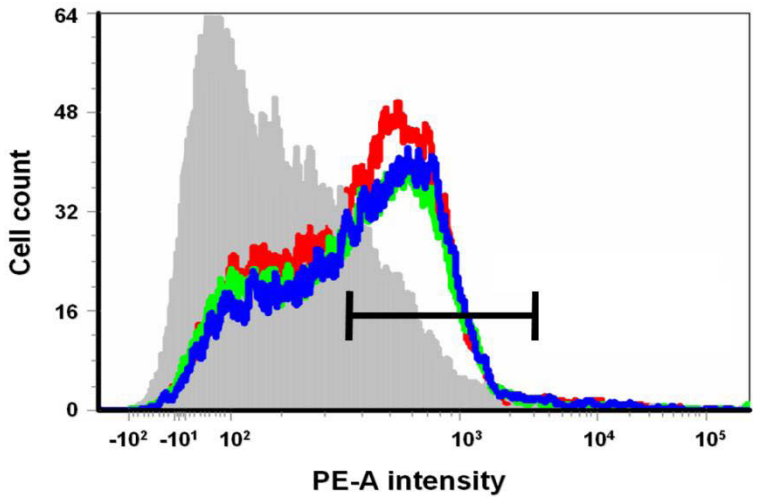

Figure 17: FACS analysis of infected *Sf*9 cells expressing the ß1,2-XylT from *A. thaliana*. (Grey) Negative control (*A.c.* NPV); (Red) *A.c.* XylT1 infected *Sf*9 cells; (Blue) *A.c.* XylT2

infected *Sf*9 cells; (Green) *A.c.* XylT3 infected *Sf*9 cells; Bar marks ß1,2-xylosylated insect cells

*Sf*9 cells infected with *A.c.* XylT1 were harvested 3 days post infection, lysed and applied to SDS/PAGE and Western Blot. Same protein amounts of ß1,2-XylT (positive control) and GFP expressing insect cells (negative control) were loaded and visualized via Coomassie staining (Figure 18; a). As shown in the Western Blot (Figure 18; b) using anti-HRP antibody, endogenous proteins from *A.c.* XylT1 infected *Sf*9 cells showed a significant higher signal as compared to GFP expressing cells. However, the background signal of the negative control resulted from core α1,3-fucosylated N-glycans typically in insects.

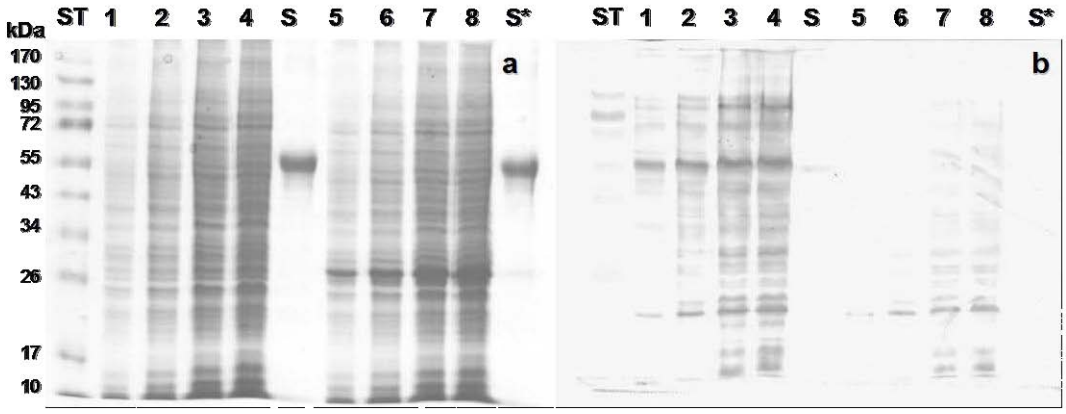

Figure 18: SDS-PAGE and Western Blot of ß1,2-XylT and GFP expressing *Sf*9 cells. (a) Coomassie blue staining; (b) anti-HRP antibody labelling; 1 – 4: ß1,2-XylT expressing insect cells in ascending amounts; S: Supernatant of ß1,2-XylT expressing insect cells; 5 - 8: GFP expressing insect cells in ascending amounts S*: Supernatant of GFP expressing insect cells

In addition to *Sf*9 lysed cells (Figure 18), the supernatant of uninfected, *A.c.* GFP and *A.c.* XylT1 infected *Sf*9 cells was applied to SDS-PAGE (Figure 19) and Western blot (Figure 19; b, c). Virus particles that budded off from ß1,2-XylT expressing *Sf*9 cells (positive control), contained ß1,2-xylosylated N-glycans on the major baculoviral envelope protein, gp64, detected by anti-HRP antibody (Figure 19; c; lane 3; lane 3*). However, gp64 of the negative control was not bound by anti-HRP antibody (Figure 19; c; lane 2; lane 2*), indicating that ß1,2-xylosylation on gp64

was a result of the recombinant plant glycosyltransferase. Of course, no gp64 in the supernatant of uninfected *Sf*9 cells was detected by anti-gp64 antibody (Figure 19; b; lane 1; lane 1*).

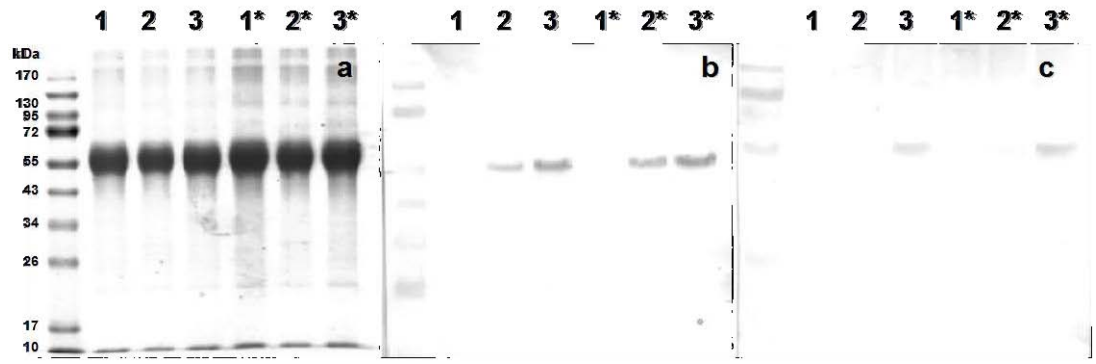

Figure 19: SDS-PAGE and Western Blot of baculoviruses budded off from ß1,2-XylT and GFP expressing insect cells. (a) Coomassie blue staining; (b) anti-gp64 antibody labelling; (c) anti-HRP antibody labelling; 1: uninfected *Sf*9 cells; 2: GFP expressing *Sf*9 cells; 3: ß1,2-XylT expressing *Sf*9 cells; * indicate double protein amount of the corresponding sample

By analyzing released monosaccharides from precipitated proteins of *A.c.* XylT1 (positive control) and *A.c.* GFP infected (negative control) *Sf*9 cells, the hypothesis of ß1,2-xylosylated N-glycan modification, as a result from the recombinant plant ß1,2-XylT, was supported. Although, the amount of GlcN, Man and Glc was almost similar between ß1,2-XylT and GFP expressing *Sf*9 cells, released Xyl was significantly higher in the positive control (Figure 20; lane B). Additionally, the amount of released Xyl from viruses that budded off from ß1,2-XylT expressing insect cells was higher in comparison to the negative control. These results supported the data shown via Western Blot (Figure 19).

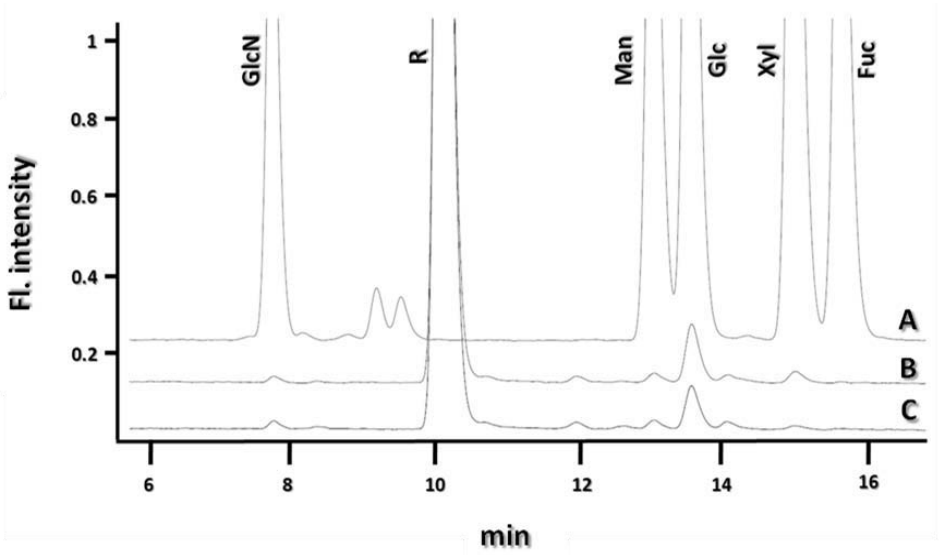

Figure 20: Monosaccharide analysis of *Sf*9 cells expressing GFP or ß1,2-XylT from *A. thaliana*. (A) Standard mix; (B) ß1,2-XylT expressing *Sf*9 cells; (C) GFP expressing *Sf*9 cells; (R) Reagent

Besides FACS analysis (Figure 17), fluorescence microscopy was used to determine ß1,2-xylosylated N-glycans of membrane proteins. *A.c.* XylT1 infected cells showed strong fluorescence at the cell surface (Figure 21; d), whereas only weak background (auto)fluorescene was detected in the case of *A.c.*NPV infected cells (Figure 21; b).

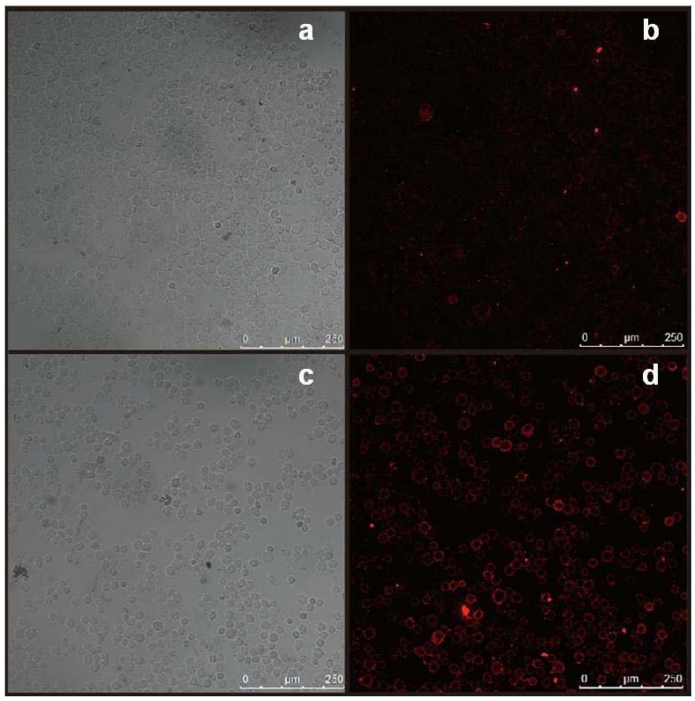

Figure 21: CLSM recording of *A.c.* NPV (negative control) and *A.c.* XylT1 (positive control) infected *Sf*9 cells. a/b: Transmission/Fluorescence (negative control) c/d: Transmission/Fluorescence (positive control)

Based on the characterization of *Sf*9 cells that expressed recombinantly a ß1,2-XylT from *A. thaliana*, a screening procedure for isolating the same enzyme from snail origin was planned. By FACS, Western blot, HPLC and fluorescence microscopy it was shown, that N-gylcans of endogenous proteins were ß1,2-xylosylated, as soon as the recombinant plant enzyme was expressed and active. These modified N-glycans were also exposed on the cell surface according to FACS and microscopy data (Figure 17; Figure 21). Thus, besides homology-based PCR and protein purification, cell sorting using FACS was performed to isolate a ß1,2-XylT gene from snail origin. For this reason, an expression cDNA library from *A. lusitanicus* was created in *Sf*9 cells and screened for those cells that exhibited similar characteristics to the positive control (*A.c.* XylT1 infected *Sf*9 cells).

4.2.2 *Screening for a ß1,2-XylT from snail origin*

4.2.2.1 Screening procedures using molecular biological tools

As shown by the recombinant expression of a plant ß1,2-XylT in *Sf*9 cells (positive control), N-glycans of membrane proteins were modified with ß1,2-Xyl linked to the ß-Man, as soon as a recombinant enzyme was expressed and active (chapter 4.2.1). Based on these data, a screening procedure for the isolation of a snail ß1,2-XylT was established. FACS analysis was used to isolate those insect cells with the highest anti-HRP antibody binding capacity from the *A. lusitanicus* expression cDNA library. These cells could contain the desired ß1,2-XylT from gastropods, based on ß1,2-xylosylated N-glycans exposed on the cell surface.

Isolated cells were mixed with wild type *Sf*9 cells to produce recombinant baculovirus that contained reduced gene diversity compared to the initial cDNA library. The virus solution produced after 8 days was used for new insect cell infection. Those cells that bound most anti-HRP were isolated again. Baculovirus DNA was isolated of each cell sorting round and its gene diversity was tested by control PCR. As shown in Figure 22 there was no significant correlation between the number of cell sortings and loss of gene diversity. Actually, the diversity between the initial cDNA library and the baculovirus resulting from 3 succeeding cell sorting rounds were similar (Figure 22; lane 1 vs. lane 4). In addition, no single gene could be amplified during the cell sorting process (Figure 22; lane 2 - 4).

Isolation of cells from the *A. lusitanicus* expression cDNA library based on the anti-HRP epitope by MACS did not result either in single gene amplification nor in decrease of gene diversity (data not shown).

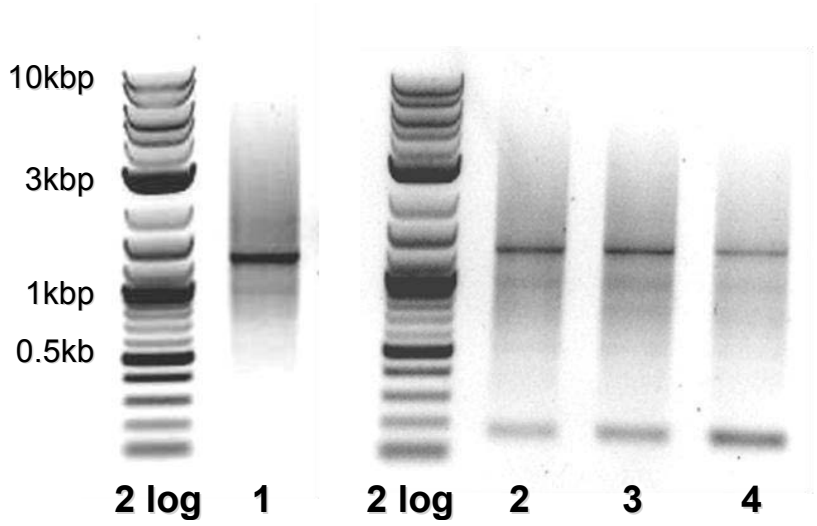

Figure 22: Gene diversity from baculoviruses that resulted from several cell sorting rounds using FACS analysis. (1) Initial cDNA library; (2) Gene diversity of the first cell sorting; (3) Gene diversity of the second cell sorting; (4) Gene diversity of the third cell sorting

In addition to the screening procedure by FACS, homology-based PCR was performed to isolate a ß1,2-XylT gene from gastropods. For this reason, degenerated primers specific for highly conserved nucleotide sequences within *Arabidopsis thaliana* (AJ272121.1), *Hordeum vulgare* (AM179853.1), *Populus alba* (AJ891042.1) and *Solanum tuberosum* (AM179855.1) ß1,2-XylT genes were designed (5'GATTYGAGTAYGCAAAYCTCTTYCA3' and 5'GTGWGTBAGWCCWGCBCCAT GAGC3'). Screening the created cDNA libraries from *A. lusitanicus, A. fulica, B. glabrata* and Bge cells did not result in an amplification of the corresponding ß1,2-XylT fragment (data not shown). Although several PCR conditions were tried, such as varying the annealing temperature, de-/increasing the Mg^{++} concentration, synthesis of novel cDNA or changing the template amount, the reactions were not successful in amplifying a fragment of a snail ß1,2-XylT.

PCRs using alternative primers, e.g. specific to the conserved amino acid sequence KP/AWPI/R, did not result in the amplification of a correct gene fragment. With these new forward primers, several PCR conditions were tried: different templates (cDNA library and non-ligated double-stranded cDNA), annealing temperatures ranging from 50°C to 66°C, several reverse primers and variations of magnesium concentrations. In summary, it was not possible to amplify a desired fragment of a snail ß1,2-XylT.

4.2.2.2 Protein purification

4.2.2.2.1 N-glycans of microsome proteins from A. lusitanicus are ß1,2-xylosylated or α1,3 fucosylated, respectively

Besides the use of molecular biological tools (chapter 4.2.2.1) to find the ß1,2-XylT gene from gastropods, first steps of ß1,2-XylT protein purification from isolated snail microsomes were performed. In addition to previous studies that detected ß1,2-xylosylated N-glycans from different snail species (Gutternigg et al, 2004; Gutternigg et al, 2007), in this work microsomes of *A. lusitanicus* were tested for ß1,2-xylosylated proteins.

Isolated *A. lusitanicus* and rabbit brain microsomes, respectively, were spottet on a nitrocellulose membrane and tested by anti-HRP antibody. Fetuin was used as the negative control, because it lacks ß1,2-xylosylated and α1,3-fucosylated N-glycans (Figure 23; -). Bromelain, which was used as the positive control, was bound strongly by anti-HRP antibody (Figure 23; +). Rabbit microsomes showed almost no signal, even when 3µg protein were spottet (Figure 23; A2), indicating that ß1,2-xylosylated and α1,3-fucosylated N-glycans were not present. On the other hand, snail microsomes were bound strongly by anti-HRP antibody (Figure 23; B1, B2), which confirmed that mollusc proteins are ß1,2-xylosylated and/or α1,3 fucosylated.

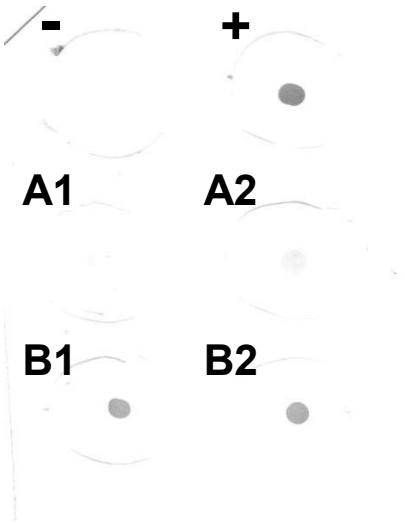

Figure 23: Dot blot of isolated rabbit and snail (*A. lusitanicus*) microsomes using anti-HRP antibody. (-) Fetuin (2µg); (+) Bromelain (2µg); (A1) 1.5µg rabbit microsomes; (A2) 3µg rabbit microsomes; (B1) 1.5µg snail microsomes (*A. lusitanicus*); (B2) 3µg snail micorsomes (*A. lusitanicus*)

Before the purification of the ß1,2-XylT from homogenized snail tissue and/or from isolated microsomes could be started, it was tried to establish an enzyme activity assay. Either pure recombinant ß1,2-XylT from *A. thaliana* (gift from Univ.-Prof. Dr. Lukas Mach) or snail microsomes were used as the enzyme source. Several acceptor glycans, with or without a fluorescent label (PA) were tested (MM-PA, TAKA-GP (M^5/M^6), GnGn-OS, GnGnF^6-OS, GnGn-PA; Figure 24).

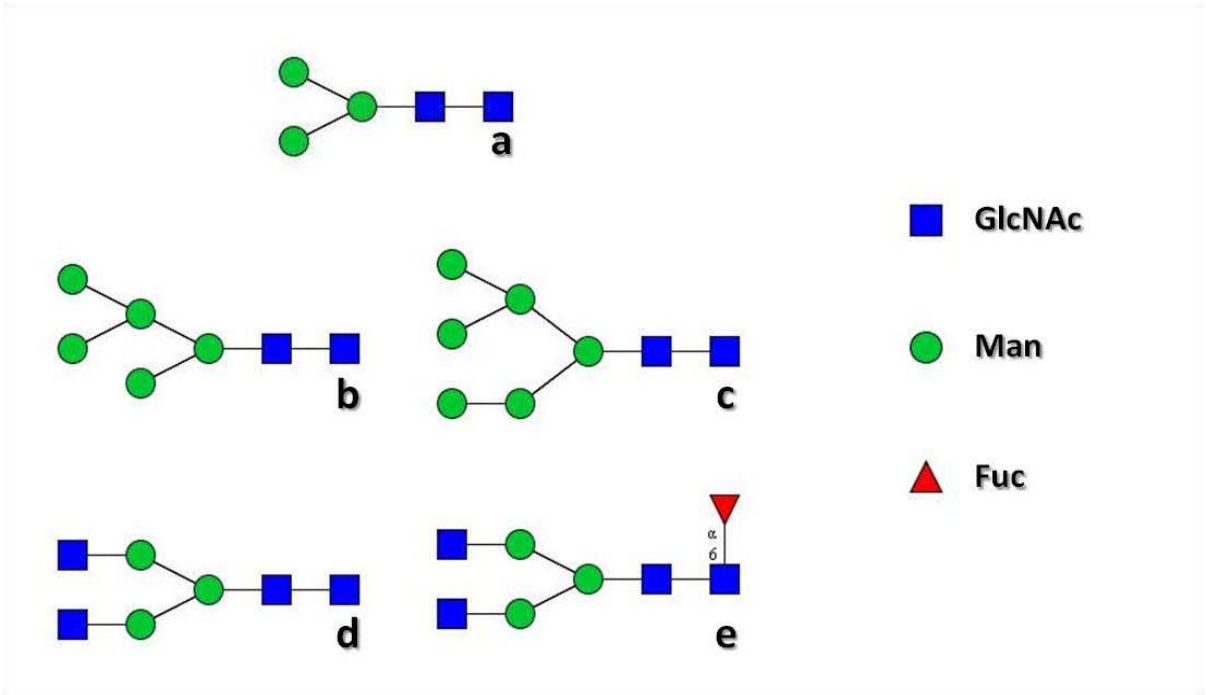

Figure 24: Acceptor glycan structures used for testing ß1,2-XylT activity. (a) MM (b) M^5 (c) M^6 (d) GnGn (e) GnGnF^6

GnGn-PA was the best acceptor glycan, although the signal intensity was low (data not shown). Based on this information, GnGn-PA was used as acceptor substrate for all further assays. The acceptor substrate was purified from fibrin (data not shown). As shown in Figure 25, in total 6 enzyme activity assays with isolated *A. lusitanicus* microsomes were performed. The number of enzyme activity assays corresponded to different microsome preparations. The first five showed a clear tendency by having higher signal intensities when the acceptor substrate was added to the mixture, indicating that GnGn-PA was modified by ß1,2-Xyl. The result of the sixth assay was opposite, so that the negative control without the acceptor substrate had higher signal than the sample. Using HPLC as the analytical tool, it was not possible to detect any ß1,2-xylosylated GnGn-PA. Based on these results, that showed low

enzyme activity and not a clear reproducibility, purification of a ß1,2-XylT from homogenized snail tissue and/or isolated microsomes could not be performed.

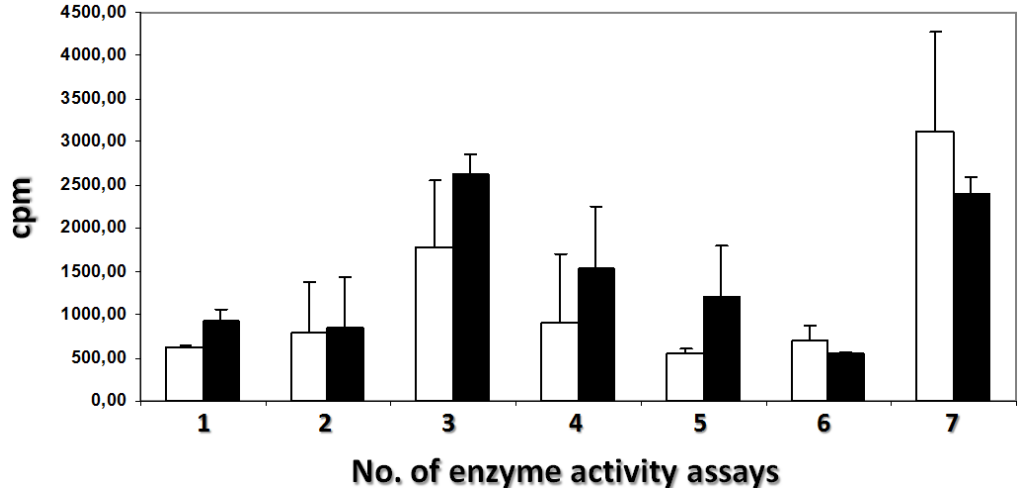

Figure 25: ß1,2-XylT activity assays by using snail microsomes as the enzyme source. Enzyme activity assays performed on (1) 2.3.11 (2) 15.3.11 (3) 18.3.11 (4) 29.3.11 (5) 8.4.11 (6) 14.4.11; 3h incubation (7) 14.4.11; 20h incubation

4.3 ppGalNAcT

4.3.1 *Isolation of the full-length ppGalNAcT gene*

*A full-length cDNA library from Bge cells was created in order to isolate and characterize the first mollusk ppGalNAcT. The PCR strategy was based on homology to the conserved regions (amino acid sequence WGGEN and VWMDEY/F) of four different ppGalNAcTs (Homo sapiens 2 NP_004472, Mus musculus NP_038842, Caenorhabditis elegans NP_498722 and Drosophila melanogaster AAQ56699). By using degenerated primers an 187bp product was amplified. In order to receive the full length ppGalNAcT gene, an inverse PCR of self-ligated cDNA was performed, which resulted in the amplification of a 1,659bp product coding for 553 amino acids (6 –558). The remaining 5' (amino acid 0 – 417) and 3' end (amino acid 481 – 600) was amplified by RACE-PCR (Figure 26).**

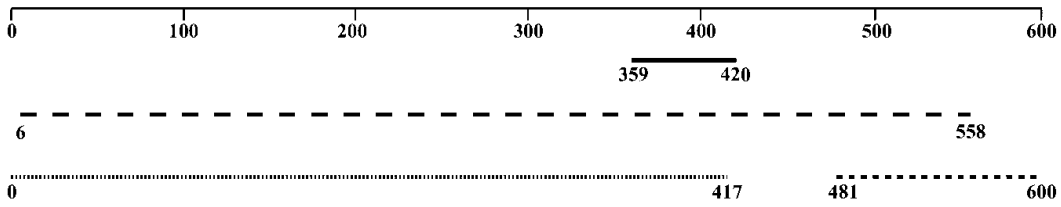

Figure 26: Strategy for isolating specific fragments of the novel ppGalNAcT gene. (359 – 420): homology-based PCR; (6 – 558): inverse PCR of self-ligated cDNA; (0 – 417): 5' RACE PCR; (481 – 600): 3' RACE PCR

4.3.2 *ppGalNAcT gene analysis*

Thereby, the full-length ppGalNAcT gene could be recovered, which turned out to be coding for a 600 amino acid type-II membrane protein containing a putative N-terminal cytoplasmic tail, a transmembrane domain (amino acid 7–25, predicted by TMHMM Server v. 2.0), a stem region, a luminal catalytic domain and a ricin-like motif at the C-terminus similar to almost all of the ppGalNAcTs. An amino acid sequence alignment with mammalian and invertebrate ppGalNAcT enzymes is shown in Figure 27. Highly conserved motifs, such as the DxH sequence followed by two

*reduced cysteine residues ($D^{252}{}_{S}HC_{E}C^{257}$) and the $D^{353}{}_{LKMDV}WGGEN_{L}E^{365}$ were identified within the snail glycosyltransferase too.**

```
H.sapiens           -------MRRRSRMLLCFAFLWVLGIAYYMYSGGGS-----ALAGGAGGGAGRKEDWNEI 48
M.musculus          -----MRKFAYCKVVLATSLVWVLLDMFLLLYFSEC---NKCE----------------- 35
C.elegans           MLSVGGGRSAVCRAVIATSIVWLLIDVVILFYYLDP---STSQQQPFPEDNRILNRARRI 57
D.melanogaster      -------MLPRFRSFYGKLIIFILVALCFILYSKVQ--------QNGSPEEPPVAPLVRA 45
C.gigas             ------MPRRRCRWVLFLGLVWAVLLIMYVYRSPGD-----KVDGNGSPNHYGDAVREMV 49
B.glabrata          ------MVRRKLRLLVILAGIWLVGIVVYLFKGDDQSEPEKRVIDKGPDGFYVHKEEKEH 54

H.sapiens           DPIKKK-----DLHHSNGEEKAQSMETLP-----------------PGKVRWPDFNQEAY 86
M.musculus          --------------------EKQERGLP------AGDVLELVQKPHEGP---GEMGKPVV 66
C.elegans           EPLPPAAQHDSDPDAHPIQPEKQEKQVYPVDKETANQLRKLMETQAFGPGYHGQGGTGVT 117
D.melanogaster      AALRGHGRERFEAYSDSENEIARPATQSP-------YEQIIQLDLQKQKVGLGEQGVAVH 98
C.gigas             VPEQQE-EQAIPKRIPKKAEHQQEVRKEE------------PDEEAGDIPWEEFDEKGY 95
B.glabrata          QPFQEENKAEVLHQVEDQWKKKQDAEITQSRQTFIETKKLLPPDDELGDIPWGQFDELGY 114

H.sapiens           VGGTMVRSGQDPYARNKFNQVESDKLRMDRAIPDTRHDQCQR---KQWRVDLPATSVVIT 143
M.musculus          IPKEDQEKMKEMFKINQFNLMASEMIALNRSLPDVRLEGCKT---KVYPDNLPTTSVVIV 123
C.elegans           VPEDKKTIKEKRFLENQFNVVASEMISVNRTLPDYRSDACRTSGNNLKTAGMPKTSIIIV 177
D.melanogaster      LSGAAKERGDEIYKKIALNEELSEQLTYNRSVGDHRNPLCAK--QRFDSESLPTASVVII 156
C.gigas             IDKKRCSAGQDCYNRNKFNQLASDNTKSNRHVPDTRNAQCRE---EQHDSNLDPTSVIIT 152
B.glabrata          ISKTTLKPGQDPYARNKFNLQASDNIKSNRHVPDTRHMNCRS---ETWSQDLPDTSVIIT 171
                                  *   *          * *   *               *  *

H.sapiens           FHNEARSALLRTVVSVLKKSPPHLIKEIILVDDYSN-----DPEDGALLGKIEK--VRVL 196
M.musculus          FHNEAWSTLLRTVHSVINRSPRHMIEEIVLVDDASERDFLKRPLESYVKKLKVP--VHVI 181
C.elegans           FHNEAWTTLLRTLHSVINRSPRHLLEEIILVDDKSDRDYLVKPLDSYIKMFPIP--IHLV 235
D.melanogaster      FFNEPYSVLLRTVHSTLSTCNEKALKEIILVDDGSDNVELGAKLDYYVRTRIPSGKVTIL 216
C.gigas             FHNEARSTLLRTIVSVFSRSPKHLITEIILVDDFSD-----DPSDGQELAVIKR--VKIL 205
B.glabrata          FHNEARSALLRTIVSIFRKSPDHLIREIILVDDFSD-----DPSDGQELDKIKK--VKVL 224
                    * **    ****  *           ** **** *

H.sapiens           RNDRREGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVI 256
M.musculus          RMEQRSGLIRARLKGAAVSRGQVITFLDAHCECTAGWLEPLLARIKHDRRTVVCPIIDVI 241
C.elegans           HLENRSGLIRARLTGSEMAKGKILLFLDAHVEVTDGWLEPLVSRVAEDRKRVVAPIIDVI 295
D.melanogaster      RLKNRLGLIRARLAGARIATGDVLIFLDAHCEGNIGWCEPLLQRIKESRTSVLVPIIDVI 276
C.gigas             RNDKREGLMRSRVKGADAARAPILTFLDSHCECNVGWLEPLLDRIKGDRTRVVSPIIDVI 265
B.glabrata          RNDKRQGLIRSRVNGANMAKGKVLTFLDSHCECNEKWLEPLLDRVKQDRRNVVSPIIDVI 284
                        * ** * *  *          *** * *    * ***  *    *  *  ******

H.sapiens           NMDNFQYVG--ASADLKGGFDWNLVFKWDYMTPEQRRS-----RQGNPVAPIKTPMIAGG 309
M.musculus          SDDTFEYMA--GSDMTYGGFNWKLNFRWYPVPQREMDR-----RKGDRTLPVRTPTMAGG 294
C.elegans           SDDTFEYVT--ASETTWGGFNWHLNFRWYAVPKRELNR-----RGSDRSMPIQTPTIAGG 348
D.melanogaster      DANDFQYSTNGYKSFQVGGFQWNGHFDWINLPEREKQRQRRECKQEREICPAYSPTMAGG 336
C.gigas             NMDNFEYIG--ASADLKGGFDWNLVFKWDYMTPEERNK-----RAGNPIQPIRTPMIAGG 318
B.glabrata          SMDNFDYIG--ASADLKGGFDWNLVFKWDYMSAEERNR-----QRQNPTAPIRTPMIAGG 337
                        * *          *** *   * *                       *   *  ***

H.sapiens           LFVMDKFYFEELGKYDMMMDVWGGENLEISFRVWQCGGSLEIIPCSRVGHVFRKQHPYTF 369
M.musculus          LFSIDRDYFQEIGTYDAGMDIWGGENLEISFRIWQCGGTLEIVTCSHVGHVFRKATPYTF 354
C.elegans           LFAIDKQFFYDIGSYDEGMQVWGGENLEISFRVWMCGGSLEIHPCSRVGHVFRKQTPYTF 408
D.melanogaster      LFAIDRRYFWEVGSYDEQMDGWGGENLEMSFRIWQCGGTIETIPCSRVGHIFRDFHPYKF 396
C.gigas             LFSIEKKWFEELGKYDRNMDVWGGENLEISFRVWQCHGSLEIIPCSRVGHVFRKQHPYTF 378
B.glabrata          LFSIDKSWFDELGQYDLKMDVWGGENLEISFRVWQCHGNLEIIPCSRVGHVFRKQHPYTF 397
                    **      *   * **  *  ******* *** * * *  *   ** *** **   ** *

H.sapiens           PGGSGTVFARNTRRAAEVWMDEYKNFYYAAVP-SARNVPYGNIQSRLELRKKLSCKPFKW 428
M.musculus          PGGTGQIINKNNRRLAEVWMDEFKNFFYIISP-GVTKVDYGDISSRLGLRRKLQCKPFSW 413
C.elegans           PGGTAKVIHHNAARTAEVWMDEYKAFFYKMVP-AARNVEAGDVSERKKLRETLQCKSFKW 467
D.melanogaster      PNDR-DTHGINTARMALVWMDEYINIFFLNRPDLKFHADIGDVTHRVMLRKKLRCKSFEW 455
C.gigas             PGGSGNVFARNTRRAAEVWMDNYKEFYYAAVP-SAKMVNFGDISERMDLRKRLSCKPFKW 437
B.glabrata          PGGSGQIFARNTKRAAEVWMDEYIQFYFAAVP-SAKHVDVGDISERLALRDRLQCKPFKW 456
                    *         *  * * ****          *        *    *  **  * ** * *

H.sapiens           YLENVYPELRVPDHQDIAFGALQ--QGTNCLDTLGHFADG--VVGVYECHNAGG-NQEWA 483
M.musculus          YLENIYPDSQIP-RHYFSLGEIRNVETNQCLDNMARKENE--KVGIFNCHGMGG-NQVFS 469
C.elegans           YLENIYPEAPLP-ADFRSLGAIVNRFTEKCVDTNGKKDGQ--APGIQACHGAGG-NQAWS 523
D.melanogaster      YLKNIYPEKFVPTKDVQGWGKVHAVNSNICLDDLLQNNEKPYNAGLYPCGKVLQKSQLFS 515
C.gigas             FLEHVYPELKVPGHQDQAFGSIQ--QDNNCMDTLGNFADG--ILGIFPCHFAGG-NQEFS 492
B.glabrata          FLENVYPELKIPSVQDIAFGSIK--QGNDCMDTMGHFADG--ILGLYPCHNSGG-NQEFS 511
                     *   **    *       *         * *           *   *       *

H.sapiens           LTKEKSVKHMDLCLT-VVDRAPGSLIKLQGCRENDS--RQKWEQIEGNSKLRHVGSNLCL 540
M.musculus          YTANKEIRTDDLCLD---VSKLNGPVTMLKCHHLKG--NQLWEYDPVKLTLQHVNSNQCL 524
C.elegans           LTGKGEIRSDDLCLSSGHVYQIGSELKLERCSVSKINVKHVFVFDDQAGTLLHKKTGKCV 583
D.melanogaster      FTNTNVLRNELSCATVQHSESPPYRVVMVPCMENDE---FNEQWRYEHQHIIHSNTGMCL 572
C.gigas             LTKEGFIRHLDLCVT-LTGSMPGTVVKLFQCQEGNT--LQMWQRTSRDTMLKSHNYDLCL 549
B.glabrata          LTKAGEVKHLDLCVT-LVDTRPGNEVKLYQCTPGNY--KQQFVQNPAKDQLRHKSYDLCL 568
                     *          *                  *                        *

H.sapiens           DSRTAK-SGGLSVEVCGP-ALSQQWKFTLNLQQ--- 571
M.musculus          DKATEEDSQVPSIRDC-TGSRSQQWLLRNVTLPEIF 559
C.elegans           TGADQR----VTLDECGLGRKDQMWQLEGYQSP--- 612
D.melanogaster      DHQGLKSLDDAQVAPCDPHSESQRWTIEH------- 601
C.gigas             DSQEVQ-MKGLIVNACQEGSPTQKWSFAVSKGG--- 581
B.glabrata          DSVVWQ-TKGIVANKCDPSSYTQKWTFSLSKNR--- 600
                           *      * *
```

Figure 27: Multiple sequence alignment (ClustalW) of known ppGalNAcTs with the Bge enzyme: *Homo sapiens* 2 (NP_004472), *Mus musculus* (NP_038842), *Caenorhabditis*

elegans GLY-3 (NP_498722), *Drosophila melanogaster* (AAQ56699), *Crassostrea gigas* (EKC38600) and *Biomphalaria glabrata* (KC182513). Identical amino acids are marked with asterisks

*By establishing a phylogenetic tree based on the neighbour joining method, it was hypothesized that the snail enzyme is a ppGalNAcT-2 isoform. Compared to all 20 human enzymes, the Bge ppGalNAcT was closest related to the human ppGalNAcT-2 isoform (data not shown). In comparison to the ppGalNAcTs belonging to the group Ib classified by (Bennett et al, 2012) the hypothesis was strengthen, because all type 2 isoforms (Drosophila melanogaster T2 (NP_608773.2), Danio rerio T2 (NP_001121823.1), Xenopus tropicalis T2 (XP_002931524.1), Gallus gallus T2 (XP_419581.2), Homo sapiens T2 (X85019), Crassostrea gigas T2 (EKC38600), Caenorhabditis elegans GLY4 (NP_507850.2)), including the snail ppGalNAcT, have the same root within the tree (Figure 28). The closest relationship (63% identity) was to the ppGalNAcT-2 from C. gigas.**

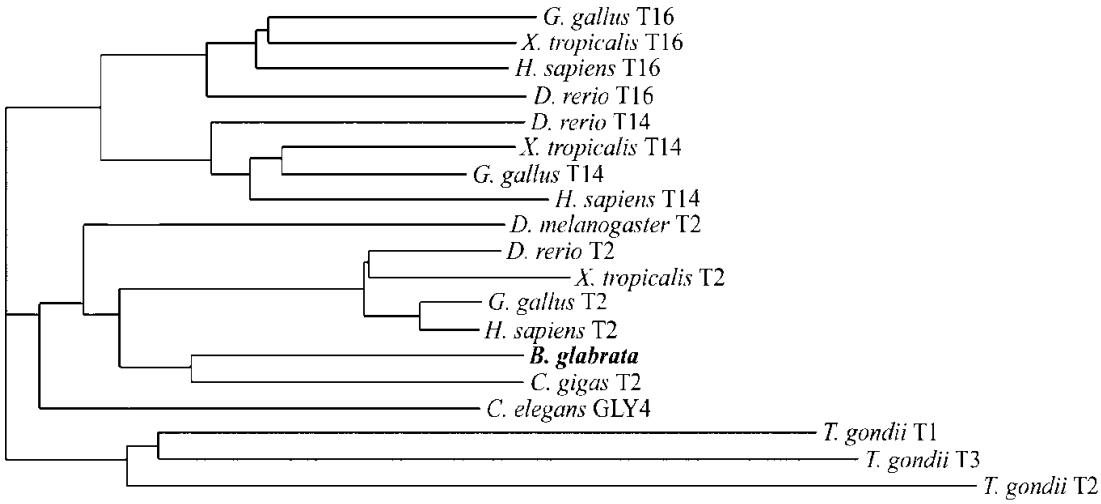

Figure 28: Branch (group Ib according to (Bennett et al, 2012)) of the phylogenetic tree of ppGalNAcTs. The phylogenetic tree is based on the neighbor-joining method based on amino acid sequences using ClustalW software. The branch length represents evolutionary distance between the members of the ppGalNAcT family. The following Genbank accession numbers were used: *Gallus gallus* T16 (XP_001231965.1), *Xenopus tropicalis* T16(NP_001039091.1), *Homo sapiens* T16 (AJ505951), *Danio rerio* T16 (XP_001339749.3), *Danio rerio* T14 (NP_001038460.1), *Xenopus tropicalis* T14 (NP_001072369), *Gallus gallus* T14 (XM_419370.2), *Homo sapiens* T14 (Y09324), *Drosophila melanogaster* T2 (NP_608773.2), *Danio rerio* T2 (NP_001121823.1), *Xenopus*

tropicalis T2 (XP_002931524.1), *Gallus gallus* T2 (XP_419581.2), *Homo sapiens* T2 (X85019), *Biomphalaria glabrata* (KC182513), *Crassostrea gigas* T2 (EKC38600), *Caenorhabditis elegans* GLY4 (NP_507850.2), *Toxoplasma gondii* T1 (XP_002365147.1), *Toxoplasma gondii* T3 (XP_002369811.1), *Toxoplasma gondii* T2 (XP_002365091.1)

4.3.3 *Biochemical characterization of the novel ppGalNAcT from B. glabrata*

4.3.3.1 Purification and determination of ppGalNAcT specificity

*A truncated version of the Bge ppGalNAcT without the transmembrane domain was expressed in Sf9 cells. The supernatant was tested for ppGalNAcT enzyme activity, as well as the supernatant of Sf9 cells expressing another glycosyltransferase. Whereas the supernatant of cells infected with ppGalNAcT expressing baculovirus showed high enzyme activity, no transfer at all was detected in the control supernatant, confirming no interfering ppGalNAcT activity derived from the baculovirus insect cell expression system. Recombinant ppGalNAcT was further purified using metal chelate affinity chromatography. Quality of the purification was determined by Coomassie staining and immunoblotting using anti-His antibody. A band at approximately 65 kDa was detected, which correlated with the molecular weight as calculated by the sum of the amino acids of the truncated form (Figure 29).**

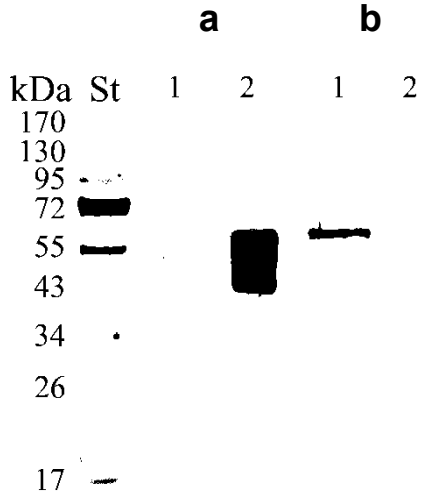

Figure 29: Purification control of recombinant snail ppGalNAcT by SDS-PAGE and Western blot analysis. (a) Coomassie staining; (b) Immunoblotting with Penta-His monoclonal antibody; 1: purified ppGalNAcT; 2: Fetuin, as the negative control

For further characterization, the purified sample (specific activity 46 mU/ml; one unit is defined as the amount of enzyme that transfers 1 µmol of GalNAc in 1min of

the standard reaction mixture) was subjected to enzyme activity assays. First, four different peptides (Muc1a, Muc1a' Muc2 and Muc5Ac) were chosen as acceptors for enzyme activity. Qualitative analysis was done by MALDI-TOF MS; for quantification of the substrate conversion the relative amount of peak areas of the HPLC patterns were used for calculation. All four arbitrarily selected peptides were functional acceptor substrates, but with different sugar transfer efficiency. By an incubation time of 45min, almost 95% of Muc2 peptide (PTTTPITTTTTVTPTPTPTGTQTK) was converted to a monoglycosylated state. In the case of Muc5Ac (GTTPSPVPTTSTTSAP) about 36%, for Muc1a (APPAHGVTSAPDTRPAPGC) 31% and for Muc1a' (AHGVTSAPDTR) 6% were monoglycosylated (Figure 30), indicating that the primary sequence plays an important role for being used as good/weak acceptor substrate.

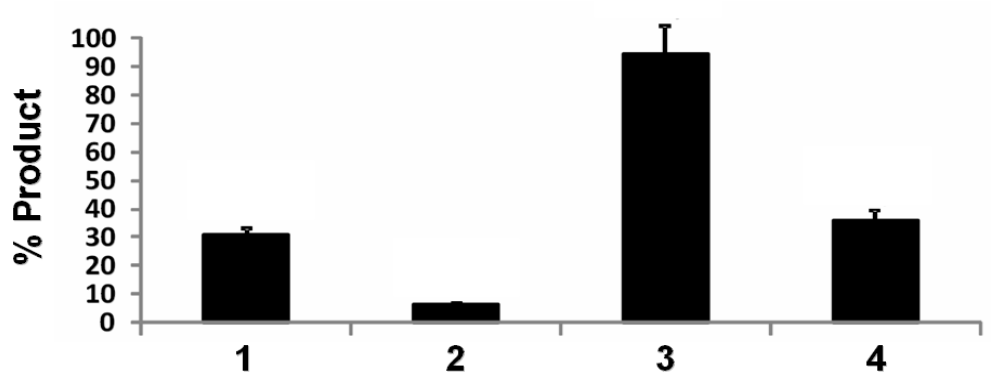

Figure 30: Conversion of different acceptor substrates to a monoglycosylated state under standard conditions. (1) Muc1a; (2) Muc1a'; (3) Muc2; (4) Muc5Ac

Furthermore, the first glycosylation site during an incubation time of 45min was tried to determine by mass spectrometry (ETD reagent). For Muc2 peptide (T_{15}) and Muc1a (T_8) it was possible to identify a single Thr as the O-glycosylation site. By coupling a TMT-label (Thermo Scientific) to Muc5Ac an additional charge was added, resulting in triple charged peptide. By this procedure, the fragmentation was improved, so S_5 could be identified as the O-glycosylation site. T_5, as well S_6 could be the glycosylated aa in the case of Muc1a'. In this case it was not possible to detect a single O-glycosylation site.

Similar to already described ppGalNAcTs, we observed the transfer of multiple GalNAc residues upon extended incubation times. Muc2 turned out to be the best acceptor where the addition of a second GalNAc residue was detectable already after

90min and clearly visible after 120min (Figure 31; d). Figure 32 illustrates the conversion of Muc2 to a diglycosylated state measured by HPLC. By increasing the incubation time up to 24h it was possible to detect a transfer up to 8 GalNAc residues onto Muc2 acceptor peptide (Figure 31; h).

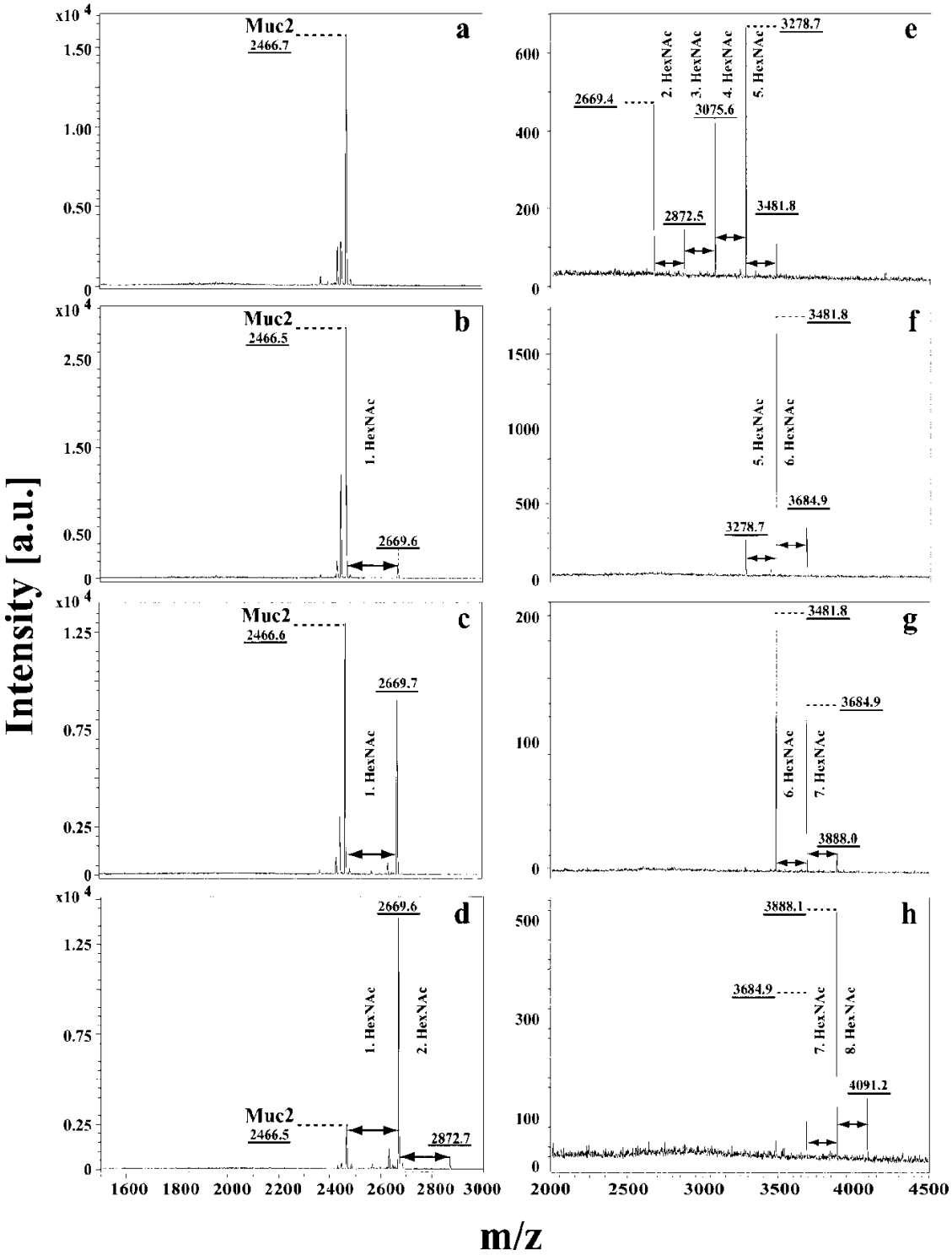

Figure 31: Mass spectrometric ([M+Na]+) analysis of the transfer of GalNAc using Muc2 as acceptor peptide. (a) Muc2 peptide; (b) 10min incubation; (c) 60min; (d) 2h; (e) 3h; (f) 6h; (g) 9h (h) 24h

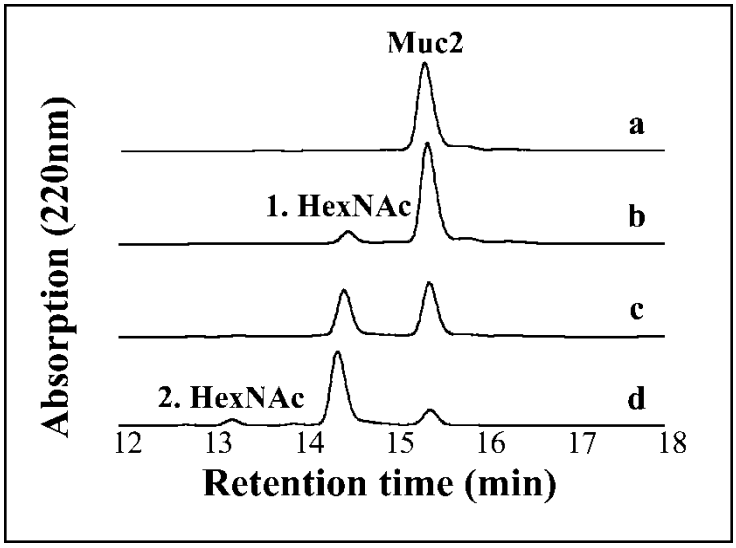

Figure 32: HPLC analysis of the transfer of GalNAc onto Muc2 acceptor peptide. (a) Muc2 peptide; (b) 10min incubation; (c) 60min; (d) 2h

Figure 33 illustrates the glycosylation sites of 8 GalNAc residues linked to Muc2, which was determined by mass spectrometry. It is conspicuous that at least one amino acid is between two neighbouring glycosylation sites.

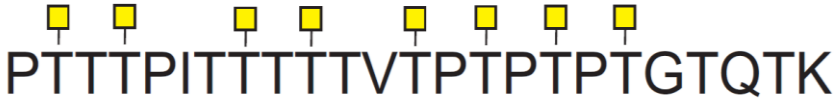

Figure 33: Detection of 8 O-glycosylation sites of Muc2 peptide after an incubation time of 24h

On Muc5Ac, which was the second best acceptor peptide, traces of a second sugar residue could be seen after 2h (Figure 34; c), but less than 50% of Muc1a or Muc1a' were converted into monoglycosylated peptides within 2h.

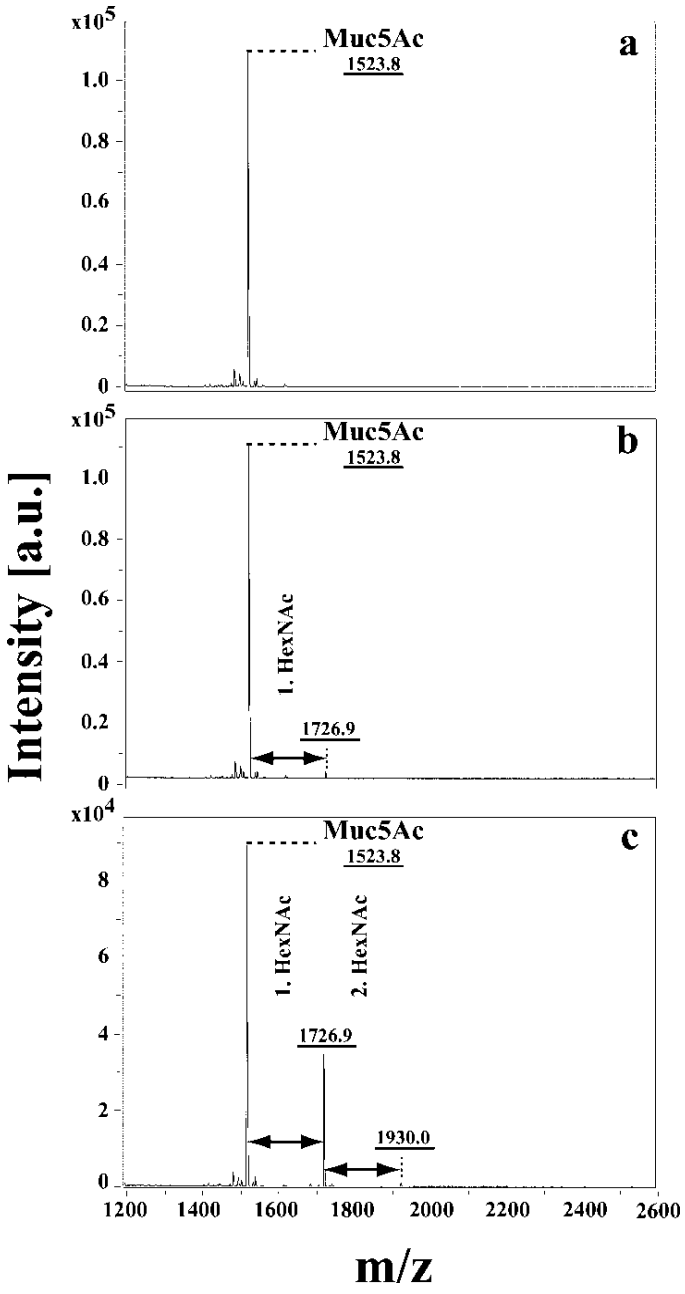

Figure 34: Mass spectrometric ([M+Na]+) analysis of the transfer of GalNAc onto Muc5Ac acceptor peptide. (a) Muc5Ac peptide; (b) 20min incubation; (c) 120min

Based on the identification of T_{15} as the first glycosylation site in Muc2, as well as on data described in (Gerken et al, 2006), 22 peptides (Table 8; Table 9) were designed to determine the influence of neighbouring amino acid relative to the O-

glycosylation site (-3 to +3). As a first experiment, a peptide containing the optimal core region for human ppGalNAcT-2 (Gerken et al, 2006), CHT1, was substituted with aa preferred from human ppGalNAcT-1. The conversion to the monoglycosylated state of "CHT1" within 45min, was assumed as 100%. The other tested peptide conversions (CHT2 – CHT15), were related to CHT1.

The second experiment was the other way around by designing a peptide, CHT16, that contained an optimal core region for human ppGalNAcT-1. CHT17 – CHT22 contained changed amino acids of the CHT16 core, that were described as being more preferred from human ppGalNAcT-2 (Gerken et al, 2006). The product conversion of CHT17 – CHT22 were calculated in relation to CHT16, which was assumed as 100% under standard conditions.

Table 8: Peptides used for determination of the detailed acceptor substrate specificity. Red highlighted amino acids were substituted within the core region of CHT1

Amino acid sequence	**Peptide name**	**% product conversion (in relation to CHT1)**
APPAHPGPTPGPRPAPG	CHT1	100
APPAHPGVTPGPRPAPG	CHT2	26.4
APPAHPGPTPGYRPAPG	CHT3	104.6
APPAHPGPTPGKRPAPG	CHT4	0
APPAHPGPTPGHRPAPG	CHT5	12.9
APPAHPGVTPGYRPAPG	CHT6	3.7
APPAHPGITPGPRPAPG	CHT7	0
APPAHPGFTPGPRPAPG	CHT8	0
APPAHPGPTEGPRPAPG	CHT9	86.8
APPAHPGPTIGPRPAPG	CHT10	29.8
APPAHPGPTPRPRPAPG	CHT11	26.4
APPAHHGPTPGPRPAPG	CHT12	110.8
APPAHPVPTPGPRPAPG	CHT13	41.2
APPAHPGLTPGPRPAPG	CHT14	0
APPAHHVVTERYRPAPG	CHT15	0

Table 9: Peptides used for determination of the detailed acceptor substrate specificity. Green highlighted amino acids were substituted within the core region of CHT16

Amino acid sequence	Peptide name	% product conversion (in relation to CHT16)
APPAHDFVTPAPRPAPG	CHT16	100
APPAHDFPTPAPRPAPG	CHT17	91.4
APPAHPFVTPAPRPAPG	CHT18	133
APPAHIFVTPAPRPAPG	CHT19	113
APPAHVFVTPAPRPAPG	CHT20	115
APPAHDGVTPAPRPAPG	CHT21	87.8
APPAHPGPTPAPRPAPG	CHT22	120

The substitution of Pro at position -1 in the core region of CHT1 with four different hydrophobic amino acids, resulted in a very strong decrease of the monoglycosylated product (Val: -74%; Ile: -100%; Phe: -100%; Leu: -100%). When Pro at position +1 was substituted with Glu or Ile, a decrease of -13% and -70%, respectively, was detected. Pro at position +3 was changed to Lys (-100%), His (-87%) or Tyr. Very surprisingly, Tyr enhanced the conversion up to 5% in comparison to CHT1 peptide. Another not expected enhancement factor was measured after the substitution of Pro -3 with His. It resulted in almost 11% more conversion into the product. The Gly residues at -2 and +2 were changed to Val (-59%) and Arg (-74%), respectively (Table 8; Figure 35). Some combinations were also tested in the case of CHT6 and CHT15, which reduced the conversion to almost 100% in both cases (Table 8).

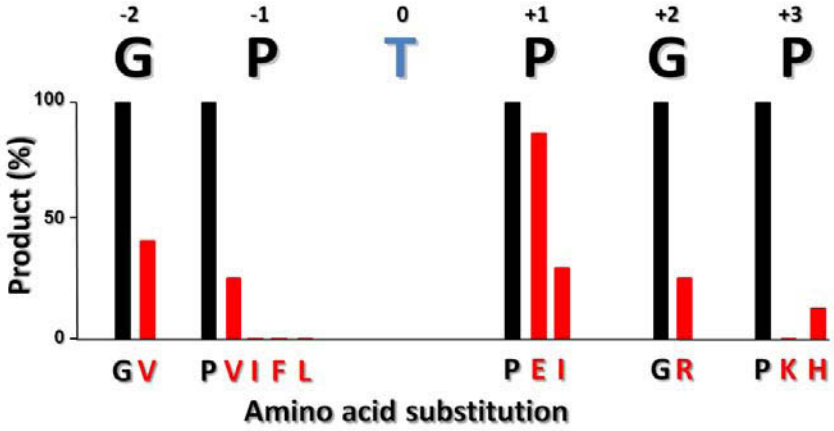

Figure 35: Influence of neighbouring amino acids within the core region of CHT1, relative to the O-glycosylation site, in terms of product conversion

Only 16% monoglycosylated CHT16, an optimal peptide acceptor for human ppGalNAcT-1, was produced from Bge ppGalNAcT under standard conditions. In comparison, CHT1, containg the optimal core for human ppGalNAcT-2, was converted to 67% product under same conditions. The change of Val at position -1 to Pro within the CHT16 core, did not result in an enhanced product conversion (-8.6%), as expected from CHT2 conversion. In this case, Pro at position -1 from CHT1 core was substituted with Val and resulted in a strong product reduction (-74%) indicating that Pro at -1 is strongly preferred from the snail ppGalNAcT (Table 8; Figure 35). On the other hand, as expected from data described in (Gerken et al, 2006), the substitution of Asp at position -3 with Pro (+33%), Ile (+13%) and Val (+15%) resulted in higher product conversion. Gly instead of Phe at position -2 resulted in ~12% reduction, whereas the substitution of "DFV" (-3 to -1) within the CHT16 core with "PGP" enhanced product conversion up to 20% (Table 9; Figure 36).

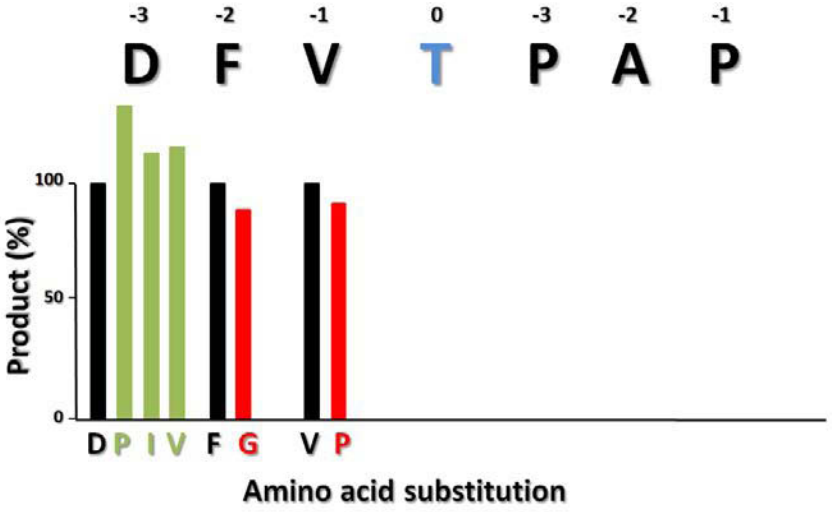

Figure 36: Influence of neighbouring amino acids within the core region of CHT16, relative to the O-glycosylation site, in terms of product conversion

4.3.3.2 Glycosylation of the enzyme

According to the amino acid sequence no N-glycosylation site was expected, however, O-glycans linked to the recombinant enzyme expressed in insect cells could be detected by lectin blot analysis using peanut agglutinin (specific for the T-antigen (Galß1-3GalNAcα1-Ser/Thr)) in combination with O-glycosidase (specific for the same O-glycan) treatment (Figure 37).

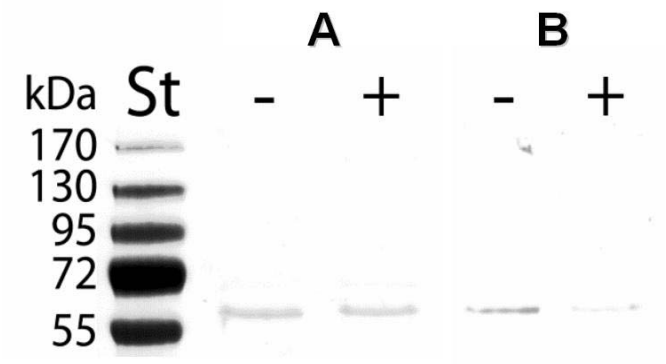

Figure 37: Determination of O-glycosylation by SDS-PAGE and Western blot analysis of purified ppGalNAcT. (A) Coomassie staining before (-) and after (+) O-glycosidase incubation. (B) Lectin blot with peanut agglutinin before (-) and after (+) O-glycosidase incubation.

4.3.3.3 Properties of the enzyme

*The activity of the recombinant ppGalNAcT was not affected by storage for 72h in a temperature range from −20°C to room temperature, or by the addition of up to 10% of MeOH or glycerol. Addition of 5% of AcCN had no negative effect on enzyme activity, whereas 10% of AcCN or lyophilization reduced enzyme activity to about 60%. The activity was drastically reduced at storage temperatures above room temperature. After 72h at 37°C less than 10% of activity was detectable. However, 37°C was the optimal incubation temperature for short assays up to 2h (data not shown). Further, ppGalNAcT showed a pH-optimum at 6.0–6.5 using MES as the appropriate buffer salt (Figure 38; a). The enzymatic transfer was dependent on divalent cations (no activity in the presence of EDTA) with the order of increasing support being Mn^{2+}>Co^{2+}>Mg^{2+}>Ca^{2+}. Cu^{2+}-ions abolished activity completely (Figure 38; b). Standard assays resulting in nearly 100% of conversion into product were also carried out with half amount of enzyme to confirm the validity of the original assay. Maximal rates of transfer were achieved with Mn^{2+}-concentrations from 10–20mM. Concentrations above 80mM reduced the enzyme activity (data not shown). Analysis of enzyme activity revealed a K_M of 0.064mM, with v_{max} at 0,16nmol/min/µg for Muc2 acceptor peptide and a K_M of 0.046mM and v_{max} at 0,054nmol/min/µg for UDP-GalNAc in the presence of constant amounts of Muc2.**

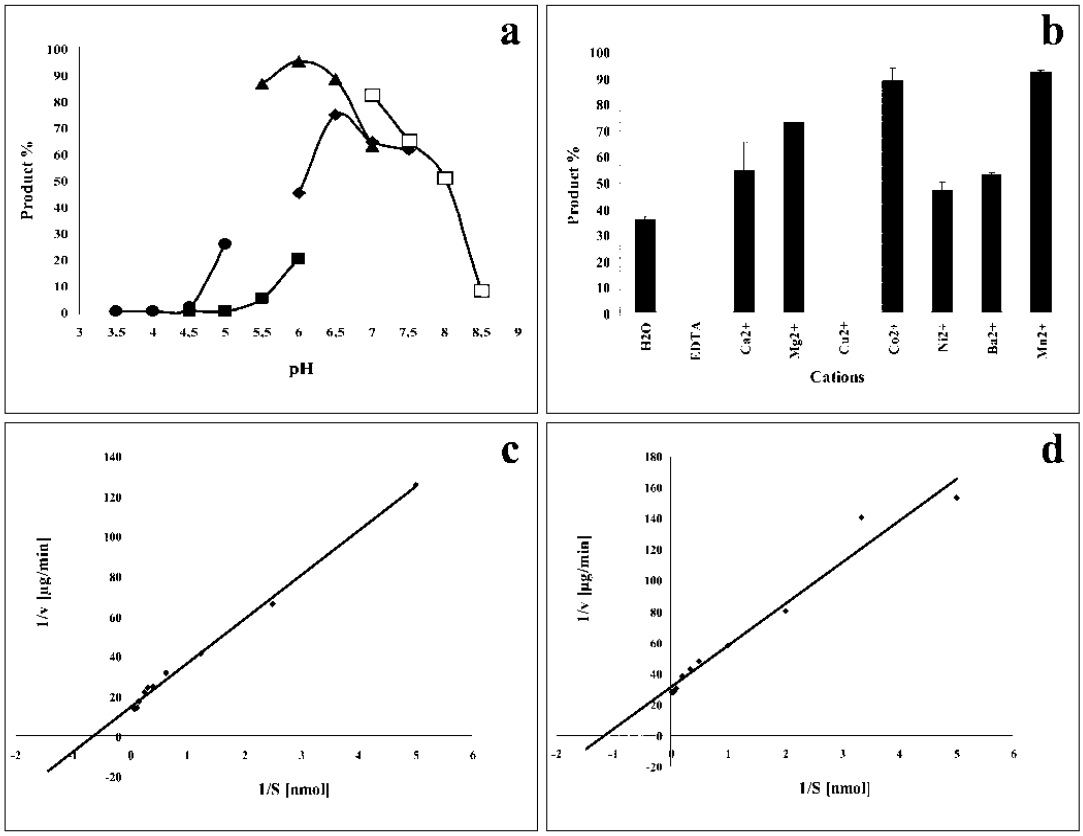

Figure 38: Properties of ppGalNAcT from Bge cell line. a) pH pattern using the following buffer systems: (● - ●) acetate/NaOH, (■ - ■) citrate/NaOH, (▲ - ▲) MES, (♦ - ♦) phosphate and (□ - □) Tris/HCl; b) Influence of divalent cations on enzyme activity; c) Determination of K_M and V_{max} (Lineweaver Burk plot) for Muc2 acceptor peptide; d) Determination of K_M and V_{max} (Lineweaver Burk plot) for UDP-GalNAc substrate

The ppGalNAcT from *B. glabrata* has a high donor substrate specificity, because it transfers only GalNAc from UDP-GalNAc to Muc2 and Muc5Ac, respectively, but not Gal (UDP-Gal), Fuc (GDP-Fuc), Xyl (UDP-Xyl), GlcNAc (UDP-GlcNAc) or NANA (CMP-NANA) from the respective activated nucleotides. Additionally, these sugars were not transferred to the GalNAc moiety linked to Muc2, indicating that the novel enzyme has a strict specificity for Tn-antigen synthesis *in vivo*.

---

* Results were described in Taus et al, 2013

4.4 T-synthase

4.4.1 *Characterization of the isolated N-terminus*

Based on highly conserved amino acid residues within T-Synthases from *Homo sapiens* (AAF81981.1), *Mus musculus* (AAF81982.1), *Caenorhabditis elegans* (AAG36940.1) and *Drosophila melanogaster* (AAF52724.1) successful homology-based PCRs were performed from all tested cDNA libraries (Figure 39). A 189bp fragment coding for 63 amino acids could be further isolated from Bge cells, which showed 75% identity to *C. gigas* (EKC27137.1), 65% to *M. musculus* (AAH64767.1), 62% to *H. sapiens* (NP_064541.1).

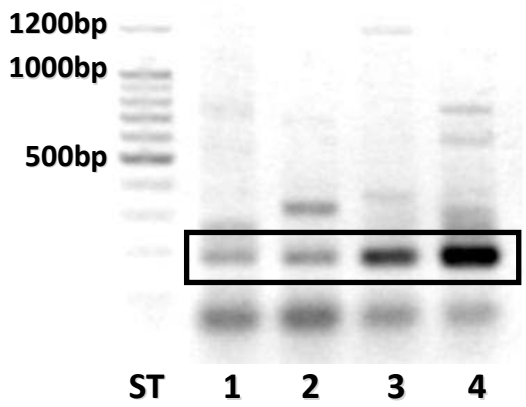

Figure 39: Homology-based PCR for amplifying a T-synthase fragment from four different cDNA libraries. (1) *A. lusitanicus*; (2) *A. fulica*; (3) *B. glabrata*; (4) Bge cell line; the black box marks the T-synthase fragment

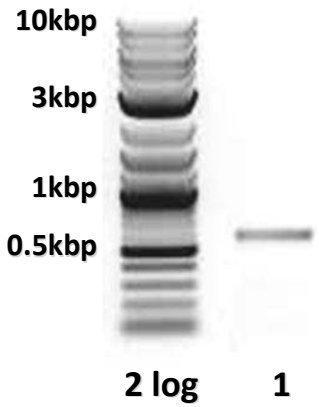

Figure 40: Amplification of the 5' T-synthase end by RACE PCR. (1) Amplified PCR fragment coding for the T-synthase 5' end from Bge cells

A nested 5'RACE PCR resulted in the amplification of an about 600bp fragment (Figure 40). This fragment contained an open reading frame coding for only 58 amino acids. Uploading this fragment to the *Biomphalaria glabrata* genome project (http://biology.unm.edu/biomphalaria-genome/) Contig6173.4 was identified that contained the coding region of the 5' T-synthase end. An alignment of Contig6173.4 with the amplified fragment from 5'RACE PCR, showed a missense mutation leading to a wrong stop codon after 58 amino acids. By correcting this mutation an open reading frame of 173 amino acids was obtained, which had high homology to other T-synthases. The 5' end of the snail T-synthase showed 55% to *M. musculus*, 52% to *H. sapiens* and 51% identity to *C. gigas*. An N-terminal transmembrane domain, ranging from aa 7 to 29, was identified using TMHMM Server, v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). An alignment of the N-terminal end from *B. glabrata*, *H. sapiens* (NP_064541.1), *M. musculus* (AAH64767.1) and *C. gigas* (EKC27137.1) is shown in Figure 41. The detection of conserved $D^{81}DD$ and $T^{116}WG$ motifs, which are typical of ß1,3-glycosyltransferases (Malissard et al, 2002) supported the possibility for the identification of the first core 1 ß1,3-GalT from gastropods.

Primers specific to various parts in the known 5' T-synthase gene were tested to isolate the missing 3' end using RACE PCRs. These primers were also used in several combinations (nested PCR) to allow higher specificity and the amplification from very low template amount. Up to now, several amplified fragments were sequenced, but no correct coding region for the complete 3'end was identified, indicating that PCR conditions done so far were too unspecific.

```
H.sapiens       --------------------MASKSWLNFLTFLCGSAIGFLLCS---------------- 24
M.musculus      --------------------MASKSWLNFLVFLCGSAIGFFLCS---------------- 24
C.gigas         MATRKDNKEKVTKRGKVLGCVQTSHLNFCFGLSVGLVISFLIASFSSLQTPLILTRKYAA 60
B.glabrata      --------------------MNAPPKQFLLGLLAGLSLTFLIYSYVHLS--------YRE 32
                                    : :      : :  *  : *:: *

H.sapiens       QLFSILLGEKVDTQPNVLHNDPHARHSDDNGQNHLEGQMNFNADSSQHKDENTDIAENLY 84
M.musculus      QLLSILLREEAAIQPNMLHNDPHARHSDDNGHSHLKGQMNFNADSSQHKDENIDVAENLY 84
C.gigas         NQFDSYLSSRSLQEVAETYEDMHEEMDKRLSKNEVK-EVLFEDD-HKHHDDDA-VARKLA 117
B.glabrata      PPHPTEFRSTSLLDSLSKKQDSHEQFDQVIARSALP-EVHLDEDKHMHHDDDK-EAKQLA 90
                      : .    :     :* * . ..  .:. :  :: :: *   *:*::   *.:*

H.sapiens       QKVRILCWVMTGPQNLEKKAKHVKATWAQRCNKVLFMSSEENKDFPAVGLKTKEGRDQLY 144
M.musculus      QKVKILCWVMTSPQNLEKKAKHVKATWAQRCNKVLFMSSEENQDFPTVGLKTKEGREQLY 144
C.gigas         EEIRILCWVMTGPQNLDKKAIHVKKTWGKRCTKLIFFSSVTNNTFPTIGLNVSEGREHLT 177
B.glabrata      EKIKVLVWVMTNPNNLEKKAKIVKETWGKRCNKVIFFSSETNNDFPTIGLDVPEGREHLT 150
                ::::::* ****.*:**:***  ** **.:**.*::*:**  *: **::**.. ***::*

H.sapiens       WKTIKAFQYVHEHYLEDADWFLKADDDTYVILDNLRWLLSKYDPEEPIYFGRRFKPYVKQ 204
M.musculus      WKTIKAFQYVHDHYLEDADWFMKADDDTYVIVDNLRWLLSKYDPEQPIYFGRRFKPYVKQ 204
C.gigas         GKTMQAFKYVHDNFFDEADWFMKADDDTYFIMENLRYFLSSQDKMEPVYFGHHFKTIVRQ 237
B.glabrata      AKTMRGFRYVYEHHFNDADWFMK---------------------------------- 173
                 **::.*:**:::.:::****:*

H.sapiens       GYMSGGAGYVLSKEALKRFVDAFKTDK-CTHSSSIEDLALGRCMEIMNVEAGDSRDTIGK 263
M.musculus      GYMSGGAGYVLSKEALRRFVNAFKTEK-CTHSSSIEDLALGRCMEIINVEAGDSRDTIGK 263
C.gigas         GYYSGGAGYILSKETLRRLATTGQDPKFCRQDGGAEDAELGKCMQNLGVRTANSTDALGR 297
B.glabrata      ------------------------------------------------------------

H.sapiens       ETFHPFVPEHHLIKGYLPRTFWYWNYNYYPPVEGPGCCSDLAVSFHYVDSTTMYELEYLV 323
M.musculus      ETFHPFVPEHHLIKGYLPKTFWYWNYNYYPPIEGPGCCSDIAVSFHYVDGTTMYELEYLV 323
C.gigas         SRFHCFDPETHLMGGYPN---WYYKYDANGAKKGLGSISDYAISFHYVGPRKMYGLEFFI 354
B.glabrata      ------------------------------------------------------------

H.sapiens       YHLRPYG-------------------YLYRYQPTLPERILKEISQANKNEDTKVKLGNP-- 363
M.musculus      YRLRPYG-------------------CLYRYQPALPENILKEINQVNRKEDTKIKLGNP-- 363
C.gigas         YHLRPYGISNGLQNLNWPGLESISDYSVSFHYIKPESMYSLEFFVYHLRPYGIVVGNQDL 414
B.glabrata      ------------------------------------------------------------

H.sapiens       -------------------
M.musculus      -------------------
C.gigas         NRRVTSYTKNATQSDKVKN 433
B.glabrata      -------------------
```

Figure 41: Alignment of T-synthases from *H. sapiens* (NP_064541.1), *M. musculus* (AAH64767.1), *C. gigas* (EKC27137.1) and the N-terminal end of *B. glabrata.* * marks conserved amino acid residues

4.5 C-type lectin from *A. lusitanicus*

4.5.1 *Sequence characterization*

30 colonies containing the pBacPAK8 cDNA library from *A. lusitanicus* were sequenced to check the gene diversity. Besides parts of a putative 3-oxoacyl carrier protein from *P. aeruginosa*, an unknown gene, an N1-acetyltransferase, ubiquitin and an elongation factor, we identified the carbohydrate recognition domain (CRD) of a C-type lectin. The almost 18kDa protein showed 46%, 39% and 45% identity to Incilarin A, B and C. These proteins represent C-type lectins that were isolated from the body surface mucus of *Incilaria fruhstorferi*, which is another snail species (Yuasa et al, 1998). Highly conserved cysteine residues within the CRD of C-type lectins from *H. sapiens* (1HUP_A), *M. musculus* (EDL41704.1), *Tropidechis carinatus* (D2YVJ8.1), *Incilaria fruhstorferi* (BAA19861.1) and from *A. lusitanicus* are indicated with gray bars (Figure 42). The highly conserved sugar binding motifs EPN and WND were present in the C-type lectin from *A. lusitanicus*, indicating a Man binding specificity (Figure 42; red bars)

```
H.sapiens           ------AASERKALQTEMARIKKWLTFSLGKQVGNKFFLTNGEIMTFEKVKALCVKFQA- 53
M.musculus          ----------------------------------------------MSLDRVKALCSEFQG- 15
T.carinatus         MGRFLLVTLSLLVGAFSLNEANSCCCPQDWLPKNGFCYKVFNDHKNWNDAEMFCRKFKPG 60
I.fruhstorferi      ------MIRIVLLLVLAAQCVF-CACPNGWKEFKGYCYGFYPEKVNWLVASASCNLYGG- 52
A.lusitanicus       ------MILAVLVLSLAVTCVTSLSCQKGWLQYQENCYGFFVEPVNWLLAGASCRAQTA- 53
                                                                 .   .   *

H.sapiens           -SVATPRNAAEN----GAIQNLIK--EEAFLGITDEKTEGQFVDLTGNR--LTYTNWNEG 104
M.musculus          -SVATPRNAEEN----SAIQKVAK--DIAYLGITDVRVEGSFEDLTGNR--VRYTNWNDG 66
T.carinatus         CHLASIHSNADSADLAEYASDYLKSDGNVWIGLNDPRKQRTWVWSDRSS--TNYLAWNQG 118
I.fruhstorferi      -RLPEIDSEQRDQWLLAELTALKF--GETWAGGSARLHVGKWEWVPSLRDFSRHSHWNAG 109
A.lusitanicus       -RLTEIEDAGENQFIVSQLQKLKF--DSTWVGGSARLHIGTYEWVPTLRRVGKYTNWDQG 110
                      :.   .   .                .: * .       :           :  *: *

H.sapiens           EPNNAGSDEDCVLLLK---NGQWNDVPCSTSHLAVCEFPI--- 141
M.musculus          EPNNTGDGEDCVVILG---NGKWNDVPCSDSFLAICEFSD--- 103
T.carinatus         EPNNSKNIEYCVHLWALTGYLKWNDTPCKALYSFICQCRF--- 158
I.fruhstorferi      EPNNVANNEYCLEINQSP-AKGWNDKACTEERQFICERRVDA- 150
A.lusitanicus       EPNNYNGKEHCLELSKVV-NYRWNDFNCLENRPFVCKRKADVK 152
                    ****  . * *: :       ***  *       **:
```

Figure 42: Alignment of C-type lectin CRDs from *H. sapiens* (1HUP_A), *M. musculus* (EDL41704.1), *Tropidechis carinatus* (D2YVJ8.1), *Incilaria fruhstorferi* (BAA19861.1) and *A. lusitanicus*

4.5.2 *Recombinant expression in insect cells, E. coli and P. pastoris*

By introducing a His-tag at the C-terminal end, which was confirmed by sequencing, the recombinant expression in insect cells was checked. The expression was tested using two different insect cell lines, *Sf*9 and BTI-*Tnao*38, respectively. As shown in Figure 43, the snail C-type lectin was only expressed in BTI-*Tnao*38 cells, but not in *Sf*9. The detection with anti-His antibody was specific, because there is no band visible in the negative control (insect cells expressing GFP).

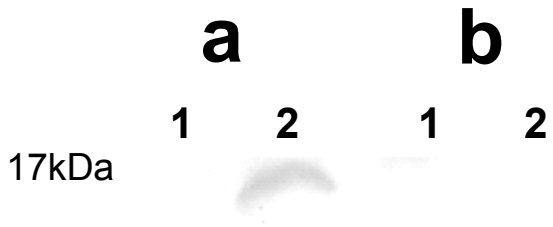

Figure 43: Recombinant expression of the C-type lectin from *A. lusitanicus* in insect cells. (a) BTI-*Tnao*38 cells; (b) *Sf*9 cells; 1: Insect cells infected with *A.c.*GFP (negative control); 2: Insect cells infected with baculovirus containing the snail C-type lectin gene

Besides insect cells, the lectin was expressed in *E. coli* and *P. pastoris*. By ligating the full-length/ΔTM gene with pET-26b for recombinant expression in bacteria, a pelB leader sequence was fused to the N-terminal end. The His-tag was fused to the C-terminus. The growth curve, which was measured for about 17h after IPTG induction (110min), indicated a strong growth inhibition when ΔTM C-type lectin was produced. The negative effect was not that clear, when the full-length gene was expressed in BL21 cells (Figure 44). The negative control was not induced with IPTG.

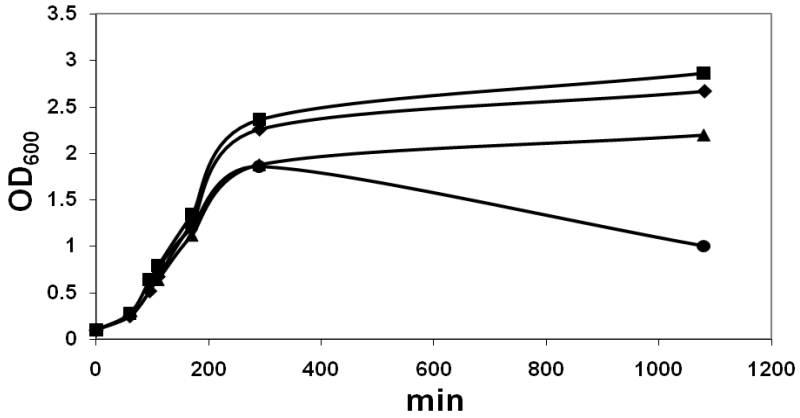

Figure 44: Growth curve of BL21 cells expressing the recombinant C-type lectin from *A. lusitanicus*. (■ - ■) negative control (full-length); (♦ - ♦) negative control (ΔTM); (▲ - ▲) full-length C-type lectin expression; (● - ●) ΔTM C-type lectin expression; induction with IPTG was initiated at 110min

The supernatant of lysed *E. coli* cells contained high amount of recombinant snail lectin (Figure 45). Based on the cleavage of the pelB leader sequence it was assumed that the main portion of recombinant lectin was located correctly in the periplasm (Figure 45; ***). A small amount was located in the cytoplasm, interpreted by a band with higher mass resulting from the still fused pelB leader (Figure 45; **). The band at about 30kDa represents a dimer (Figure 45, *), indicating that disulphide bridges could be formed in the periplasm. This result gave good indication for a correct folding of the recombinant protein. The full-length C-type lectin containing a transmembrane domain was not soluble in the supernatant of lysed bacteria cells (Figure 45; lane 1), but only detected in the cell pellet (data not shown).

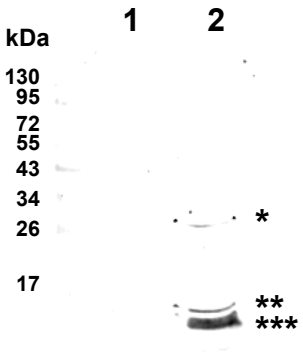

Figure 45: Recombinant expression of the C-type lectin from *A. lusitanicus* in *E. coli* cells. (1) Full-length C-type lectin (2) ΔTM C-type lectin; *: dimer; **: lectin located in the cytoplasm; ***: lectin located in the periplasm

Recombinat expression was successful in *P. pastoris* cells too. 2 tested colonies containing the full-length C-type lectin and 1 colony containing the ΔTM protein were positive concerning recombinant expression (Figure 46). In the case of yeast cells, no soluble form secreted to the supernatant could be detected (data not shown).

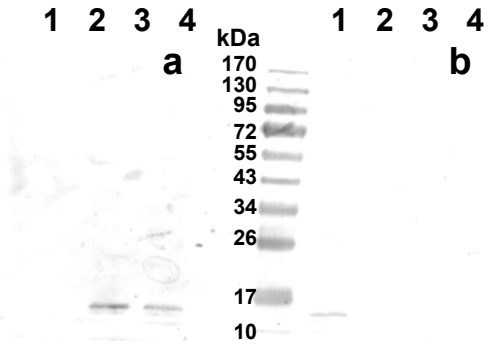

Figure 46: Recombinant expression of the C-type lectin of *A. lusitanicus* in *P. pastoris*. (a) Full-length C-type lectin; (b) ΔTM C-type lectin; 1 – 4: Single *P. pastoris* colonies

4.5.3 *Purification of the C-type lectin*

It was tried to solubilize the C-type lectin with and without detergents, before purification was performed using metal chelate affinity chromatography. Infected BTI-*Tnao*38 insect cells were treated with four different buffers and the supernatant was checked by Western blot. As shown in Figure 47, all used buffers resulted in lectin solubilization.

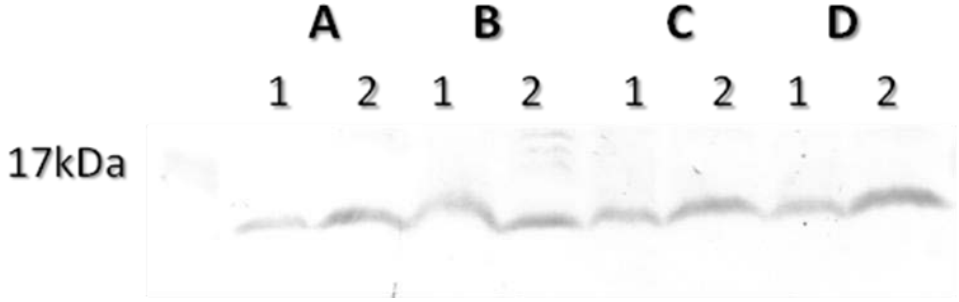

Figure 47: Solubilization of the snail C-type lectin with buffers containing different detergents. (A) 50mM Tris/HCl pH 9.0 + 1% Triton X-100; (B) 50mM Tris/HCl pH 9.0 + 1% Tween 20; (C) 50mM Tris/HCl pH 9.0 + 1% Triton X-100 + 0.5mM DTT; (D) 50mM Tris/HCl pH 9.0 + 1% Triton X-100 + 0.1% SDS; 1: Cell pellet; 2: Supernatant

In addition, infected BTI-*Tnao*38 cells expressing the C-type lectin were lysed within water, centrifuged and the supernatant checked for the presence of the lectin by Western blot. The protein was also soluble without any detergents, even if lysed cells were centrifuged at 30000xg (Figure 48).

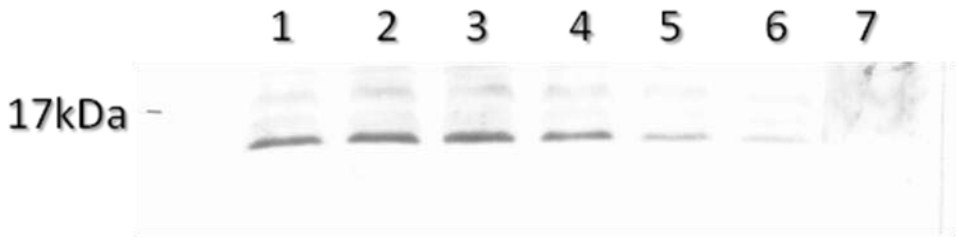

Figure 48: Solubilizing the snail C-type lectin after cell lysis (BTI-*Tnao*38) in water. Centrifugation was carried out at (1) 500xg (2) 1000xg (3) 2000xg (4) 8000xg (5) 14100xg (6) 30000xg (7) Cell pellet

After successful solubilization of the C-type lectin, protein purification was performed. In summary, no optimal purification protocol was achieved. By specific detergents, such as IGEPAL CA-630, and a reduction from 10 to 1mM imidazole within the lysis buffer, a weak binding to the Ni-NTA matrix occurred. Because the main portion of the recombinant lectin was still in the cell pellet or did not bind to the Ni-NTA matrix (flow through), the amount of eluted protein was too less for downstream processes (Figure 49).

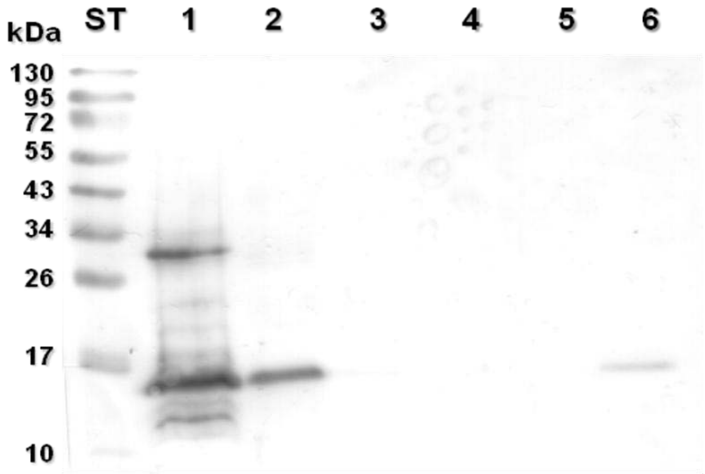

Figure 49: Purification of the C-type lectin by IGEPAL CA-630 as the detergent. (1) BTI-*Tnao*38 cell pellet; (2) Flow through; (3) Wash 1; (4) Wash 2; (5) Wash 3; (6) Elution

A better, but far away from being optimal, binding could be achieved by denaturing the C-type lectin with 8M urea. There was still recombinant protein in the flow through, but it was possible to elute higher amounts of lectin compared to the purification with detergents (Figure 50).

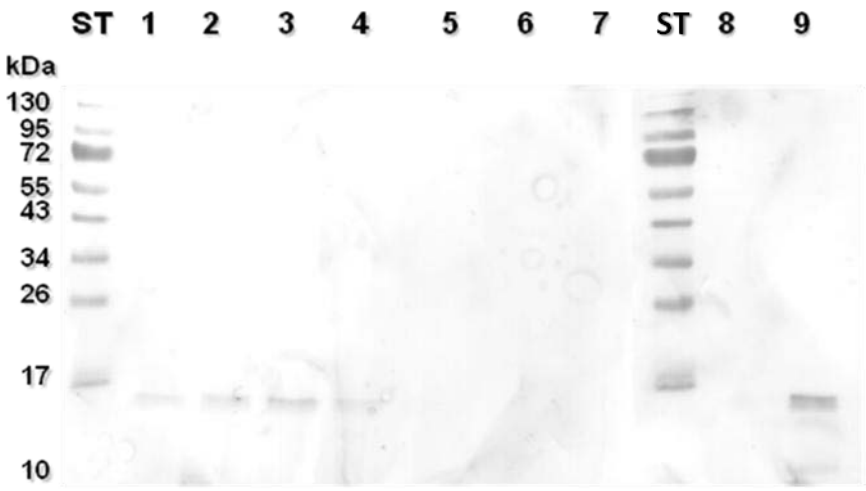

Figure 50: Purification of the C-type lectin using denaturing conditions. (1), (2), (3) Flow through; (4), (5) Buffer C; pH 6.5 (6), (7) Buffer D; pH 5.8 (8), (9) Buffer E; pH 4.5

No binding to the column occurred, when the folded C-type lectin was dissolved in buffer without detergents (data not shown). Alternatively, it was tried to purifiy the recombinant C-type lectin by mannan agarose beads. This strategy failed, because no binding to the beads was detected (data not shown).

5 DISCUSSION

5.1 Quality and future perspectives of the created cDNA libraries from snails

Several invertebrate species have been genetically studied so far and their total number of genes was estimated from different genome sequencing projects. From sequencing data, it was calculated that *C. elegans*, a free-living nematode, codes for 15000 individual genes within the 100-megabase (Mb) genome (Waterston et al, 1992; Wilson et al, 1994). *Capitella teleta*, a polychaete worm, and *Helobdella robusta*, a leech, have been sequenced and a genome size of 324Mb or 228Mb was determined, respectively. They predicted 32389 and 23400 single genes for them, where more than 50% showed orthologous with other species (Simakov et al, 2013). The haploid genome size of the oyster *C. gigas* was estimated to have approximately 637Mb. In total about 28000 genes were identified. 21085 genes matched entries in different databases, so their function could be predicted by conserved domains (Zhang et al, 2012). The owl limpet, *L. gigantea*, a sea snail, showed a genome size of 348Mb coding for 23800 predicted genes (Simakov et al, 2013). Based on the information about the number of genes from these invertebrates, it was assumed that snails are coding for 15000 to ~32000 genes, too. Thus, the aim was to create cDNA libraries from snails that contained high gene diversity.

In summary, cDNA libraries of *A. lusitanicus*, *A. fulica*, *B. glabrata* and the Bge cell line were created in *E. coli* using pSMART2IFD vector. In addition, expression cDNA libraries in *P. pastoris* (pGAPZ B) and in insect cells (pBacPAK8) of *A. lusitanicus* were prepared. Besides the goal to identify glycosyltransferase genes within these libraries, it was tried to conserve as many genes as possible within these samples for future purposes.

The pSMART2IFD cDNA libraries of all snail species showed useful gene diversity. As a representative for these cDNA libraries, several colonies containing the library of *A. lusitanicus* were sequenced. 17% of the *A. lusitanicus* cDNA library (in total $2x10^5$ colonies) contained individual genes, meaning that the library contained about 30000 single genes. Compared to the other invertebrates mentioned above, it is assumed that the cDNA library of *A. lusitanicus* contains all genes of this mollusc.

cDNA libraries of *A. fulica*, *B. glabrata* and Bge cell line, were not sequenced to determine gene diversity. However, based on more diverse inserts within these cDNA libraries compared to the one of *A. lusitanicus*, the whole transcriptome of these additional snail species should be covered, too.

The cDNA library of *A. lusitanicus* created in yeast cells, was not used for the isolation of glycosyltransferases. This sample was established to conserve as many snail genes as possible and for functional screening of alternative proteins in the future.

cDNA libraries created in this work covers the whole transcriptome approximately once. To increase the gene diversity and to achieve a higher coverage of the transcriptome, normalization methods can be used. The number of mRNA copies per gene can differ by several orders. Of course, cDNA libraries performed from mRNA represent the same proportions. To avoid this problem, normalization methods equalize the copy number of genes, so even genes transcribed at low levels can be discovered within such cDNA libraries.

5.2 Screening for a ß1,2-XylT from gastropods

The detailed neutral N-glycan profile from *A. lusitanicus* has been analyzed by (Gutternigg et al, 2004). Glycans identified in skin, viscera and eggs showed 8.6%, 5.4% and 11.5% ß1,2-xylosylated N-glycans, respectively. If almost more than 5% of all detected N-glycans show this modification, it can be suggested that the expression level, as well as the amount of glycosyltransferases is enough for isolating the gene by molecular biological methods and/or the enzyme by biochemical purification techniques. ß1,2-xylosylation was further confirmed in this study by strong binding of anti-HRP antibody to isolated *A. lusitanicus* microsomes. Protein from rabbit brain was not recognized by this antibody. To gain first information about the ß1,2-XylT from gastropods, which is an very uncommon glycosyltransferase in animals, two main strategies for its isolation were followed: 1) ß1,2-XylT gene isolation from snail cDNA libraries and 2) ß1,2-XylT protein purification from homogenized snail tissue.

To achieve a successful gene isolation, various PCR strategies and a specific cell sorting method by FACS from expression cDNA libraries were tested. Homology-based PCR utilize highly conserved domains within an enzyme family. Degenerated primers specific to these motifs can be used to amplify corresponding fragments from cDNA libraries. Once a correct fragment is sequenced, RACE-PCR leads to the full-length gene isolation. For this reason, an alignment of *Arabidopsis thaliana* (AJ272121.1), *Hordeum vulgare* (AM179853.1), *Populus alba* (AJ891042.1) and *Solanum tuberosum* (AM179855.1) ß1,2-XylT(s) genes was performed to identify conserved sequences within the catalytic domain. Although diverse PCR conditions were tried, an amplification of a corresponding ß1,2-XylT fragment from snails was not achieved. Highly conserved domains within plant ß1,2-XylTs do not seem to occur in homologous genes from gastropods.

The second method to identify a ß1,2-XylT gene from gastropods was a cell sorting process from an expression cDNA library (*A. lusitanicus*) created in insect cells. This strategy was based on the characterization of *Sf*9 cells expressing a ß1,2-XylT from *A. thaliana* (positive control). As soon as the recombinant glycosyltransferase had been expressed and was active, N-glycans of endogeneous membrane proteins were ß1,2-xylosylated. Furthermore, when the recombinant ß1,2-XylT from *A. thaliana*

was expressed in *Sf*9 cells, N-glycans of gp64, the major envelope protein of baculovirus, were modified by Xyl ß1,2-linked to the ß-Man of the N-glycan core.

With the information gained from this positive control, an expression cDNA library of *A. lusitanicus* was created in *Sf*9 cells. Insect cells were chosen for this purpose, because these host cells contain the enzymatic machinery (e.g. GlcNAcT-I) necessary to synthesize a capable acceptor glycan for the ß1,2-XylT.

It was tried to isolate those insect cells from the expression cDNA library that exposed ß1,2-xylosylated N-glycans. This modification was assumed to be the product of the snail ß1,2-XylT, as this enzyme is missing in wild type insect cells. For the isolation procedure, insect cells expressing the snail cDNA library were labeled with anti-HRP antibody. Anti-HRP antibody serum had been shown to bind complex type plant N-glycans containing Xyl and core α1,3-fucose (Kurosaka *et al*., 1991). α1,3-fucose is a common feature of insect cell N-glycans, so using this antibody resulted in some background signal. However, an increased binding of anti-HRP antibody to glycans having both, core ß1,2-xylosylation and α1,3-fucosylation, was estimated from data collected of the positive control, *Sf*9 cells expressing the plant ß1,2-XylT. FACS and MACS was used to isolate cells exhibiting the strongest antibody binding capacity. Several cell sorting rounds by FACS, as well as the attempt to enrich ß1,2-XylT expressing cells by MACS, did not yield in the amplification of the desired gene or at least in the reduction of the initial gene diversity, respectively. ß1,2-XylT expressing insect cells within the cDNA library, if at all, were in too small amounts for isolating them. Furthermore, by the binding of PE conjugated anti-HRP antibody to all insect cells exposing α1,3-fucosylated N-glycans, no enrichment of ß1,2-XylT expressing cells could be achieved using FACS. Probably an antibody only specific for ß1,2-xylosylated N-glycans would result in a more accurate cell sorting and a successful ß1,2-XylT gene isolation. Another optimization for the ß1,2-XylT gene isolation could be the creation of knock-down or knock-out insect cell mutants that lack the α1,3-FucT (EC 2.4.1.214) activity. By missing α1,3-fucosylated N-glycans, isolation based on anti-HRP antibody binding would be more specific to ß1,2-xylosylation. Up to now no insect cell line is sequenced and annotated, so the α1,3-FucT gene sequence from insect cells is unknown. Knock-down or knock-out of specific genes in insect cells requires not just knowledge of the relevant gene sequence, but also the development of gene deletion technology in insect cells. If a RNAi based knock-down of the α1,3-FucT would be

performed, the RNA sequence could be based on highly conserved domains within this enzyme family.

The last strategy was aimed to purify the ß1,2-XylT enzyme from homogenized snail tissue and/or from isolated microsomes. For this purpose, an enzyme activity assay should be established using the purified, recombinant plant ß1,2-XylT (gift from Univ.-Prof. Dr. Lukas Mach). The enzyme activity was determined by HPLC and by radioactivity. With HPLC no enzyme activity could be detected at all. In total, five enzyme activity assays using ^{3}H and ^{14}C labelled substrate, showed weak activity within snail microsomes. However, the values were too low for using this kind of detection for protein purification steps. Because HPLC analysis is much more reproducible than measuring radioactivity, in the future it should be focused on establishing assays using HPLC analysis. When several acceptor substrates were tested during this work, GnGn-PA was found to be the "best" substrate, but in general the values were very low and should be interpreted critically.

In summary, gene isolation methods, as well as protein purification strategies both have eligibility to follow in the future. An advantage of the biochemical purification is the guarantee that the purified protein exhibits native features, such as the glycosylation pattern or the folding state. Lack of enzyme activity or different enzyme variants can be the result of incorrect posttranslational modifications by recombinant expression in a not suitable host system.

A disadvantage of protein purification is the lacking information about the nucleotide sequence, once the enzyme has been isolated. However, with mass spectrometry tryptic peptides can be analyzed by LC-ESI-MS/MS with subsequent database alignment (e.g. MASCOT) to obtain at least parts of the protein sequence. Peptide sequence information from mass spectrometry of a novel enzyme can be used for designing degenerated primers and subsequent PCR reactions to isolate gene fragments from cDNA libraries. Of course, this is more specific than designing primers based on highly conserved regions from other species. Once the full-length gene is isolated, the influence of posttranslational modifications and/or specific domains can be tested by recombinant ß1,2-XylT expression in different host cells (*E. coli* vs eukaryotic expression systems) and/or deletion experiments.

5.3 Characterization of the first ppGalNAcT from *B. glabrata*

ppGalNAcTs are the largest family of glycosyltransferases, which catalyze a single linkage. This enzyme family uses UDP-GalNAc and transfers the GalNAc moiety to Ser/Thr residues of a polypeptide chain resulting in the so called Tn-antigen. They share high homology especially in the UDP-GalNAc-binding region throughout the whole animal kingdom. In snail tissues a high amount of O-glycosylation with GalNAc being the protein bound sugar has been found (Stepan et al, 2012) and thus, it could be expected to find a ppGalNAcT gene representative in gastropods. At the time the project was started, only scattered sequencing data from mollusks were available. A few Contigs were possible to find from the *B. glabrata* genome project (*Biomphalaria glabrata* genome initiative; http://biology.unm.edu/biomphalaria-genome/). In addition, parts of *Lottia gigantea* (http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html) were sequenced so far. *In the meantime, the complete sequence from Pacific oyster (Crassostrea gigas) has been published (Zhang et al, 2012) but still no expression or characterization of glycosyltransferases has been performed*.* Based on lacking sequencing data of *B. glabrata*, a full-length cDNA library was created and screened with degenerated primers, specific to highly conserved parts of ppGalNAcTs, to amplify the corresponding snail fragment by homology-based PCR.

As concluded from sequence alignments, the sequence of the amplified fragment (coding for the amino acids 359–420) showed high homology to ppGalNAcTs from several other organisms (e.g. 92% identity to Crassostrea gigas (EKC38600), 92% to Homo sapiens 2 (NP_004472), 65% to Drosophila melanogaster (AAQ56699) and 76% to Caenorhabditis elegans (NP_498722)). Three further PCR steps resulted in the complete sequence of the full length snail ppGalNAcT (600 aa). As in other ppGalNAcTs, the catalytic domain is divided into to sub-domains. The first part contained the glycosyltransferase 1 (GT1) motif (residues 166–272) described by (Hagen et al, 1999), coding for the the conserved $D^{252}xH^{254}$ sequence. This conserved motif has been shown to be involved in Mn^{2+} coordination (Fritz et al, 2004). C-terminal of the $D^{252}xH^{254}$ sequence there were two cysteine residues (255 and 257) located, which have been reported to be important for UDP-binding (Tenno et al, 2002).

The second part of the catalytic domain contained the Gal/GalNActransferase motif, which is shared among ppGalNAcTs and ß1,4-galactosyltransferases (Hagen et al, 1999). The conserved sequence, D^{353}xxxxxWGGENLE^{365}, was described to interact with the GalNAc moiety of UDP-GalNAc (Fritz et al, 2004). As usual for this enzyme family, the snail enzyme contained a C-terminal ricin type lectin domain, which consists of three homologous repeats (α, ß andγ). It has been hypothesized that this domain influences acceptor substrate preferences (Hagen et al, 1999; Hassan et al, 2000).

An extended phylogenetic tree resulted from the alignment of 102 different ppGalNAcTs (Bennett et al, 2012). According to its amino acid homology, the new snail ppGalNAcT fits into the subfamily I, group b. As expected, the enzyme from B. glabrata is closely related to the ppGalNAcT from C. gigas, which represents another mollusk. Furthermore, the snail ppGalNAcT is highly homologous to the T2-isoforms from many other species, indicating that in this work the ppGalNAcT-2 isoform from B. glabrata was isolated. Furthermore, based on similar acceptor preferences between the snail enzyme, the human ppGalNAcT-2 (Wang et al, 2003) and the insect ppGalNAc T2 (Ten Hagen et al, 2003), this hypothesis was strengthen *. The detailed acceptor substrate specificity of the snail enzyme was tested by offering various acceptor peptides. By mass spectrometric fragmentation using ETD reagent it was possible to detect the accurate O-glycosylation site within four tested peptides (Muc2, Muc5Ac, Muc1a and Muc1a'). The longer the peptides were and the more charges they exhibited, the easier was the fragmentation and the identification of the O-glycosylation site. In the case of Muc2 (24 aa) and Muc1a (19aa) the identification of a single Thr residue was determined without any problems (T_{15} and T_8, respectively). The Ser at position 5 as the O-glycosylation site of Muc5Ac (16 aa) could be identified only by coupling a label to the peptide that introduced an additional charge. In the case of Muc1a' (11 aa), the shortest peptide used for these experiments, two potential glycosylation sites (T_5 and S_6) were identified. These data indicated a strong preference for Pro at position +3 relative to the O-glycosylation site, as T_{15} for Muc2 (PTTTPITTTTTVTPTPTPTGTQTK), T_8 for Muc1a (APPAHGVTSAPDTRPAPGC) and S_5 for Muc5Ac (GTTPSPVPTTSTTSAP) were glycosylated with the first GalNAc residue. In the case of Muc2 and Muc5Ac, the O-glycosylation site was flanked by additional Pro residues at position -1 and +1, which has been described as strong enhancement factors for human ppGalNAcT-2 (Gerken

et al, 2006). Based on these data, T_5 in Muc1a' is the most likely glycosylation site, because at position +3 relative to this Thr was a Pro located (AHGVTSAPDTR). To support the assumption of a T-2 isoform isolation from Bge cells, in total 22 different peptides with specific aa substitution within the core region (-3 to +3 relative to the O-glycosylation site) were designed.

First, the product conversion of 14 peptides (CHT2 – CHT15) were calculated relatively to CHT1, APPAHPGPTPGPRPAPG, which represents the optimal core region for human ppGalNAcT-2 isoform (Gerken et al, 2006).

As expected from the data mentioned above, the substitution of Pro at -1 with Val, Ile, Phe and Leu, of Pro at +1 with Glu and Ile, as well as of Pro at +3 with Lys and His resulted in significant reduction of substrate conversion. The change of the Gly residues at position -2 and +2 to Val and Arg, respectively, influenced the conversion rate negatively as well. Furthermore, there is a Pro preference at position -3 when an optimal acceptor peptide for human ppGalNAcT-1, CHT16 (APPAHDFVTPAPRPAPG), was tested. By changing the charged aa Asp with Pro at this position, it occurred a higher product conversion of up to 33%. In addition, the substitution of "DFV" with "PGP", both from -3 to -1 relative to the O-glycosylation site, within the CHT16 core, product conversion was enhanced until 20%. Surprisingly, the aa change Val with Pro at -1 of the CHT16 core did not result in better product synthesis.

In summary, these acceptor preferences exclude the option of a ppGalNAcT-1 isoform, but strongly support the hypothesis of a T-2 isoform isolation from Bge cells (Gerken et al, 2006). Based on the missing sequencing data of the whole genome from *B. glabrata*, the total number of isoforms is not predictable. The invertebrate model systems *D. melanogaster* and *C. elegans* contain 14 and 11 members of this enzyme family (Hagen & Nehrke, 1998), respectively, so it is estimated that gastropods have also several isoforms encoded in their genome. Some of the fly isoforms are essential for normal development (Ten Hagen et al, 1998; Ten Hagen & Tran, 2002), so it may be assumed that this is the case for gastropods too.

No potential N-glycosylation site was predicted within the whole protein sequence. However, the recombinant enzyme expressed in Sf9-cells was O-glycosylated, although the native glycosylation status of this protein in the snail is not known.

Biochemical properties of the Bge ppGalNAcT correlated with those determined for other members of the family. Its stability performance was comparable, as well as

*the essential requirement of divalent cations with Mn^{2+} being optimal in a range of 10–20 mM. However, while most of the so far described ppGalNAcTss have a pH-optimum at ~7.0, the Bge ppGalNAcT shows its highest activity at pH 6.0–6.5. The K_M values are within the range of published data K_M 0.01–0.23 mM (Sørensen et al, 1995; Wandall et al, 1997) *.*

In order to gain more information about enzyme properties and the biological importance of mucin-type O-glycans, further experiments should be performed. The expression in bacteria, such as *E. coli*, would result in the lack of eukaryotic O-glycosylation of the enzyme. Using bacteria as expression system, a correlation between native glycosylation patterns and enzyme activity can be investigated. Furthemore, by deleting the C-terminal ricin like lectin domain of the snail ppGalNAcT, altered acceptor substrate specificity, especially for glycosylated peptides, can be determined.

To get an idea about the biological function of mucin-type O-glycosylation two strategies are possible. On the one hand, based on the novel gene sequence of the snail ppGalNAcT, RNAi technology could be used for inhibiting this specific enzyme activity in Bge cells. However, it can be assumed that there are more ppGalNAcT isoforms within this cell line, so probably the lack of ppGalNAcT-2 activity would be balanced by other isoforms. Of course, in this case no effect would appear. The other option is a total block of O-glycosylation. By the inhibition of the UDP-Glc/GlcNAc C_4-epimerase activity no generation of the nucleotide sugar donor, UDP-GalNAc, required for O-linked glycosylation can be synthesized. UDP-Glc/GlcNAc C_4-epimerase activity can be down regulated by RNAi based technology or by adding specific inhibitors identified by screening a uridine-based library (Winans & Bertozzi, 2002). However, GalNAc residues are not restricted to O-glycans, so a loss of UDP-GalNAc would also influence other types of glycosylation.

Most of the ppGalNAcT data presented here have been published recently (Taus et al, 2013).

---

* Discussion was described in Taus et al, 2013

5.4 T-synthase from *B. glabrata*

In the genome sequence of *L. gigantea* (http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html), ten proteins homologous to T-synthases of other species (Protein ID: 66572; 96000; 130418; 143711; 131924; 170451; 191447; 198406; 222070; 143168) could be identified. In *D. melanogaster* (http://www.ensembl.org/Drosophila _melanogaster/Info/Index) 8 genes are coding for potential core 1 ß1,3-GalTs (gene name: CG3119; CG2975; CG2983; C1GalTA; CG8708; CG34056; CG34057; tgy). Based on the fact that these representatives of invertebrates contain several isoforms of this enzyme family, different T-synthases in *B. glabrata* are assumed.

The first step for future characterization of the first T-synthase from *B. glabrata* is the isolation of the missing 3' end. Based on the T-synthase from *C. gigas*, it is predicted that about 200 aa (600 nt) are still missing from the *B. glabrata* glycosyltransferase. Optimizing RACE-PCRs give the best chance to isolate the full-length T-synthase gene. This technique is highly specific and can amplify even low expressed genes, if the conditions are optimized perfectly. Furthermore, this method resulted in successful full-length gene isolation of the novel ppGalNAcT from Bge cells after finding optimal conditions for the 3' RACE PCR. In addition, normalization methods can be performed prior to cDNA library construction to enhance the chance of full-length T-synthase gene isolation.

Once the complete gene sequence of the T-synthase is known, it should be expressed recombinantly in insect cells. Mammalian host cells should be avoided for T-synthase expression from invertebrates. As described for the worm T-synthase, it was highly active when expressed in Hi-5 cells, but only minimal activity was recovered from mammalian expressed cells (Ju et al, 2006). This indicates that invertebrate T-synthases probably need some unknown invertebrate co-factors for their function. In comparison to that, human T-synthase is not N-glycosylated in contrast to the invertebrate ones and needs Cosmc for its activity, which is a molecular chaperone necessary for functional activity (Ju & Cummings, 2002).

5.5 C-type lectin from *A. lusitanicus*

C-type lectins represent a large family of Ca^{++} dependent sugar binding proteins that share primary structural homology in their carbohydrate recognition domains (CRDs) and are found throughout the animal kingdom. These proteins, which include endocytic receptors, proteoglycans and all known collectins and selectins, are involved in many immune-system functions (Brockhausen et al, 2008). The C-type lectin domain (CTLD) was identified in many other proteins, but it is not only restricted to bind sugars. Besides sugars, CTLD binds other proteins, lipids and inorganic molecules (e.g., Ca_2CO_3). The C-type lectin fold is unique and consists of a compact domain of 110–130 amino acid residues with a double-looped, two-stranded antiparallel β-sheet formed by the amino-and carboxy-terminal residues connected by two α-helices and a three-stranded antiparallel β-sheet. However, in animals the major representatives of CTLD containing proteins are C-type lectins. Key conserved residues that bind sugars include the "EPN" and "WND" motifs within the CRD of C-type lectins. One new representative of this protein type was found in this work from *A. lusitanicus*. Based only on the primary sequence it is hard to predict the accurate binding specificity, because there are only a few contacts between the lectin and the sugar recognized. Nevertheless, some motifs give good indication for the preferred monosaccharide, especially for distinguishing Man and Gal specificity (Brockhausen et al, 2008). The C-type lectin isolated from the sea cucumber *Cucumaria echinata* contains the QPD motif (Gln-Pro-Asp) within the CRD, which is generally known to exist in Gal-specific C-type CRDs. This sequence was substituted by the motif EPN (Glu-Pro-Asn) and resulted in changed sugar binding specificity. The wild type lectin formed only complexes with GalNAc, whereas the mutant lectin containing the EPN motif was highly specific to Man (Hatakeyama et al, 2013). Based on these data and the primary sequence of the C-type lectin from *A. lusitanicus*, especially the presence of the EPN motif, a Man specificity is suggested.
Besides insect cells, during this work the snail lectin was also expressed in *E. coli* and *P. pastoris* for future purification procedures and testing its binding specificity. Up to now, no optimal purification protocol could be established for the snail C-type lectin. The recombinant protein was tried to purify with and without different detergents from the supernatant of lysed insect cells. Because it was suggested that the C-terminal His-tag without a spacer to the protein sequence was not accessible to

the Ni-NTA matrix, an N-terminal His-tag with some amino acids in between was fused in parallel. No optimal protein purification from the supernatant of lysed insect cells could be performed in both cases.

By changing the expression sytem to *E. coli* and the fusion of the C-type lectin gene with an N-terminal pelB leader, the protein was secreted to the periplasmic space to achieve correct folding by the creation of conserved disulphide bridges. While eukaryotic cells use the endoplasmic reticulum to form these bonds during the export process, in bacteria recombinant proteins are secreted to the oxidizing enviroment of the periplasm to allow a formation of disulphide bridges (de Marco, 2009). Western Blot data of the expressed C-type lectin in *E. coli* gave good indication for correct localization in the periplasm, as well as for formed disulphide bridges. The main fraction, which was the smallest version of the C-type lectin, was located within the periplasm. This truncated version results from the cleavage of the pelB leader by a signal peptidase of the cell membrane. Disulphide bridges were assumed, because dimers were also detected by Western Blot. Although these data indicated correct folding, the C-type lectin did not show any binding on mannan agarose beads.

Denaturation of the C-type lectin was performed to unfold the protein structure and to ensure better interaction between the His-residues and the Ni-NTA gel. By this strategy some binding was observed, which indicated that the His-tag had not been accessible well to Ni-NTA, but probably hidden within the core of the folded protein. A future purification under denaturing conditions would require sophisticated refolding techniques and is so far the most optimistic strategy to purify sufficient amount for testing its sugar binding specificity.

6 ABBREVIATIONS

aa	Amino acids
AcCN	Acetonitrile
AP	Alkaline phosphatase
A.c.	*Autographa californica* multicapsid nucleopolyhedro-virus
ß1,2-XylT	ß1,2-xylosyltransferase
bp	Base pair
CLSM	Confocal Laser Scanning Microscopy
Core-1 ß1,3-GalT	Core-1 ß1,3-galactosyltransferase
CRD	Carbohydrate recognition domain
FCS	Fetal calf serum
Fuc	Fucose
Gal	Galactose
GalNAc	*N*-Acetylgalactosamine
GFP	Green fluorescent protein
Glc	Glucose
GlcNAc	*N*-Acetylglucosamine
GlcNAcT-I	*N*-acetylglucosaminyltransferase I
HAc	Acetic acid
HPLC	High-performance liquid chromatography
Man	Mannose
Mb	Megabase
MeOH	Methanole
NANA	*N*-Acetylneuraminic acid
Ni-NTA	Nickel-nitrilotriacetic acid
nt	Nucleotides
ONC	Overnight culture
PA	2-aminopyridine
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PE	Phycoerythrine

ppGalNAcT	Polypeptide *N*-acetylgalactosaminyltransferase
RT	Room temperature
T-Synthase	Core-1 ß1,3-galactosyltransferase

7 LIST OF FIGURES

8 LIST OF TABLES

9 REFERENCES

Aalberse RC, Koshte V, Clemens JG (1981) Immunoglobulin E antibodies that crossreact with vegetable foods, pollen, and Hymenoptera venom. *J Allergy Clin Immunol* **68:** 356-364

Ashwell G, Harford J (1982) Carbohydrate-specific receptors of the liver. *Annu Rev Biochem* **51:** 531-554

Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA, Tabak LA (2012) Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **22:** 736-756

Brandley BK, Schnaar RL (1986) Cell-surface carbohydrates in cell recognition and response. *J Leukoc Biol* **40:** 97-111

Breg J, Van Halbeek H, Vliegenthart JF, Klein A, Lamblin G, Roussel P (1988) Primary structure of neutral oligosaccharides derived from respiratory-mucus glycoproteins of a patient suffering from bronchiectasis, determined by combination of 500-MHz 1H-NMR spectroscopy and quantitative sugar analysis. 2. Structure of 19 oligosaccharides having the GlcNAc beta(1----3)GalNAc-ol core (type 3) or the GlcNAc beta(1----3)[GlcNAc beta(1----6)]GalNAc-ol core (type 4). *Eur J Biochem* **171:** 643-654

Brockhausen I (1999) Pathways of O-glycan biosynthesis in cancer cells. *Biochim Biophys Acta* **1473:** 67-95

Brockhausen I, Schachter H, Stanley P (2008) O-GalNAc glycans. In:Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds.) Essentials of Glycobiology, Cold Spring Harbor, Laboratory Press, New York: 115–127

Brooks SA (2000) The involvement of Helix pomatia lectin (HPA) binding N-acetylgalactosamine glycans in cancer progression. *Histol Histopathol* **15:** 143-158

Burda P, Aebi M (1999) The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta* **1426:** 239-257

Burda P, te Heesen S, Brachat A, Wach A, Düsterhöft A, Aebi M (1996) Stepwise assembly of the lipid-linked oligosaccharide in the endoplasmic reticulum of Saccharomyces cerevisiae: identification of the ALG9 gene encoding a putative mannosyl transferase. *Proc Natl Acad Sci U S A* **93:** 7160-7165

Chai W, Hounsell EF, Cashmore GC, Rosankiewicz JR, Feeney J, Lawson AM (1992a) Characterisation by mass spectrometry and 1H-NMR of novel hexasaccharides among the acidic O-linked carbohydrate chains of bovine submaxillary mucin. *Eur J Biochem* **207:** 973-980

Chai WG, Hounsell EF, Cashmore GC, Rosankiewicz JR, Bauer CJ, Feeney J, Feizi T, Lawson AM (1992b) Neutral oligosaccharides of bovine submaxillary mucin. A combined mass spectrometry and 1H-NMR study. *Eur J Biochem* **203:** 257-268

de Marco A (2009) Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microb Cell Fact* **8:** 26

Dolashka-Angelova P, Beck A, Dolashki A, Beltramini M, Stevanovic S, Salvato B, Voelter W (2003) Characterization of the carbohydrate moieties of the functional unit RvH1-a of Rapana venosa haemocyanin using HPLC/electrospray ionization MS and glycosidase digestion. *Biochem J* **374:** 185-192

Drickamer K (1988) Two distinct classes of carbohydrate-recognition domains in animal lectins. *J Biol Chem* **263:** 9557-9560

Drickamer K, Dodd RB (1999) C-Type lectin-like domains in Caenorhabditis elegans: predictions from the complete genome sequence. *Glycobiology* **9:** 1357-1369

Drickamer K, Taylor ME (1993) Biology of animal lectins. *Annu Rev Cell Biol* **9:** 237-264

Duman JG, Miele RG, Liang H, Grella DK, Sim KL, Castellino FJ, Bretthauer RK (1998) O-Mannosylation of Pichia pastoris cellular and recombinant proteins. *Biotechnol Appl Biochem* **28 (Pt 1):** 39-45

Freire T, Casaravilla C, Carmona C, Osinaga E (2003a) Mucin-type O-glycosylation in Fasciola hepatica: characterisation of carcinoma-associated Tn and sialyl-Tn antigens and evaluation of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase activity. *Int J Parasitol* **33:** 47-56

Freire T, Robello C, Soulé S, Ferreira F, Osinaga E (2003b) Sialyl-Tn antigen expression and O-linked GalNAc-Thr synthesis by Trypanosoma cruzi. *Biochem Biophys Res Commun* **312:** 1309-1316

Fritz TA, Hurley JH, Trinh LB, Shiloach J, Tabak LA (2004) The beginnings of mucin biosynthesis: the crystal structure of UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-T1. *Proc Natl Acad Sci U S A* **101:** 15307-15312

Gavel Y, von Heijne G (1990) Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng* **3:** 433-442

Gerken TA, Raman J, Fritz TA, Jamison O (2006) Identification of common and unique peptide substrate preferences for the UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferases T1 and T2 derived from oriented random peptide substrates. *J Biol Chem* **281:** 32403-32416

Gerwig GJ, Kamerling JP, Vliegenthart JF, Morag E, Lamed R, Bayer EA (1992) Novel oligosaccharide constituents of the cellulase complex of Bacteroides cellulosolvens. *Eur J Biochem* **205:** 799-808

Gutternigg M, Ahrer K, Grabher-Meier H, Bürgmayr S, Staudacher E (2004) Neutral N-glycans of the gastropod Arion lusitanicus. *Eur J Biochem* **271:** 1348-1356

Gutternigg M, Bürgmayr S, Pöltl G, Rudolf J, Staudacher E (2007) Neutral N-glycan patterns of the gastropods Limax maximus, Cepaea hortensis, Planorbarius corneus, Arianta arbustorum and Achatina fulica. *Glycoconj J* **24:** 475-489

Haberl B, Kalbe M, Fuchs H, Ströbel M, Schmalfuss G, Haas W (1995) Schistosoma mansoni and S. haematobium: miracidial host-finding behaviour is stimulated by macromolecules. *Int J Parasitol* **25:** 551-560

Hagen FK, Hazes B, Raffo R, deSa D, Tabak LA (1999) Structure-function analysis of the UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. Essential residues lie in a predicted active site cleft resembling a lactose repressor fold. *J Biol Chem* **274:** 6797-6803

Hagen FK, Nehrke K (1998) cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase sequence homologs from Caenorhabditis elegans. *J Biol Chem* **273:** 8268-8277

Hall RL, Wood EJ, Kamberling JP, Gerwig GJ, Vliegenthart FG (1977) 3-O-methyl sugars as constituents of glycoproteins. Identification of 3-O-methylgalactose and 3-O-methylmannose in pulmonate gastropod haemocyanins. *Biochem J* **165:** 173-176

Hansen JE, Lund O, Engelbrecht J, Bohr H, Nielsen JO (1995) Prediction of O-glycosylation of mammalian proteins: specificity patterns of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase. *Biochem J* **308 (Pt 3):** 801-813

Harris RJ, Spellman MW (1993) O-linked fucose and other post-translational modifications unique to EGF modules. *Glycobiology* **3:** 219-224

Hart GW, Haltiwanger RS, Holt GD, Kelly WG (1989) Glycosylation in the nucleus and cytoplasm. *Annu Rev Biochem* **58:** 841-874

Hashimoto Y, Zhang S, Blissard GW (2010) Ao38, a new cell line from eggs of the black witch moth, Ascalapha odorata (Lepidoptera: Noctuidae), is permissive for AcMNPV infection and produces high levels of recombinant proteins. *BMC Biotechnol* **10:** 50

Hassan H, Reis CA, Bennett EP, Mirgorodskaya E, Roepstorff P, Hollingsworth MA, Burchell J, Taylor-Papadimitriou J, Clausen H (2000) The lectin domain of UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities. *J Biol Chem* **275:** 38197-38205

Hatakeyama T, Ishimine T, Baba T, Kimura M, Unno H, Goda S (2013) Alteration of the carbohydrate-binding specificity of a C-type lectin CEL-I mutant with an EPN carbohydrate-binding motif. *Protein Pept Lett* **20:** 796-801

Hollister J, Grabenhorst E, Nimtz M, Conradt H, Jarvis DL (2002) Engineering the protein N-glycosylation pathway in insect cells for production of biantennary, complex N-glycans. *Biochemistry* **41:** 15093-15104

Hounsell EF, Lawson AM, Feeney J, Gooi HC, Pickering NJ, Stoll MS, Lui SC, Feizi T (1985) Structural analysis of the O-glycosidically linked core-region oligosaccharides of human meconium glycoproteins which express oncofoetal antigens. *Eur J Biochem* **148:** 367-377

Hounsell EF, Lawson AM, Stoll MS, Kane DP, Cashmore GC, Carruthers RA, Feeney J, Feizi T (1989) Characterisation by mass spectrometry and 500-MHz proton nuclear magnetic resonance spectroscopy of penta- and hexasaccharide chains of human foetal gastrointestinal mucins (meconium glycoproteins). *Eur J Biochem* **186:** 597-610

Huffaker TC, Robbins PW (1983) Yeast mutants deficient in protein glycosylation. *Proc Natl Acad Sci U S A* **80:** 7466-7470

Ju T, Brewer K, D'Souza A, Cummings RD, Canfield WM (2002) Cloning and expression of human core 1 beta1,3-galactosyltransferase. *J Biol Chem* **277:** 178-186

Ju T, Cummings RD (2002) A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta 3-galactosyltransferase. *Proc Natl Acad Sci U S A* **99:** 16613-16618

Ju T, Zheng Q, Cummings RD (2006) Identification of core 1 O-glycan T-synthase from Caenorhabditis elegans. *Glycobiology* **16:** 947-958

Kishimoto T, Watanabe M, Mitsui T, Hori H (1999) Glutelin basic subunits have a mammalian mucin-type O-linked disaccharide side chain. *Arch Biochem Biophys* **370:** 271-277

Kornfeld S (1992) Structure and function of the mannose 6-phosphate/insulinlike growth factor II receptors. *Annu Rev Biochem* **61:** 307-330

Kurosaka A, Nakajima H, Funakoshi I, Matsuyama M, Nagayo T, Yamashina I (1983) Structures of the major oligosaccharides from a human rectal adenocarcinoma glycoprotein. *J Biol Chem* **258:** 11594-11598

Kurosaka A, Yano A, Itoh N, Kuroda Y, Nakagawa T, Kawasaki T (1991) The structure of a neural specific carbohydrate epitope of horseradish peroxidase recognized by anti-horseradish peroxidase antiserum. *J Biol Chem* **266:** 4168-4172

Köhler W, Prokop O, Kühnemund O (1973) Routine identification of group-C streptococci by means of an agglutinin (protectin) from the albumen gland of the edible snail, Helix pomatia. *J Med Microbiol* **6:** 127-130

Lechner J, Wieland F (1989) Structure and biosynthesis of prokaryotic glycoproteins. *Annu Rev Biochem* **58:** 173-194

Leffler H, Svanborg-Eden C (1990) Host epithelial glycoconjugates and pathogenic bacteria. *Am J Respir Cell Mol Biol* **2:** 409-411

Lis H, Sharon N (1993) Protein glycosylation. Structural and functional aspects. *Eur J Biochem* **218:** 1-27

Lommerse JP, Thomas-Oates JE, Gielens C, Préaux G, Kamerling JP, Vliegenthart JF (1997) Primary structure of 21 novel monoantennary and diantennary N-linked carbohydrate chains from alphaD-hemocyanin of Helix pomatia. *Eur J Biochem* **249:** 195-222

MacInnis AJ (1965) Responses of Schistosoma mansoni miracidia to chemical attractants. *J Parasitol* **51:** 731-746

Malissard M, Dinter A, Berger EG, Hennet T (2002) Functional assignment of motifs conserved in beta 1,3-glycosyltransferases. *Eur J Biochem* **269:** 233-239

Mason PR (1977) Stimulation of the activity of Schistosoma mansoni miracidia by snail-conditioned water. *Parasitology* **75:** 325-338

Mason PR, Fripp PJ (1977) Chemical stimulation of Schistosoma mansoni miracidial activity. *Z Parasitenkd* **53:** 287-295

Messner P, Sleytr UB (1991) Bacterial surface layer glycoproteins. *Glycobiology* **1:** 545-551

Mormeneo S, Zueco J, Iranzo M, Sentandreu R (1989) O-linked mannose composition of secreted invertase of Saccharomyces cerevisiae. *FEMS Microbiol Lett* **48:** 271-274

Mulder H, Dideberg F, Schachter H, Spronk BA, De Jong-Brink M, Kamerling JP, Vliegenthart JF (1995a) In the biosynthesis of N-glycans in connective tissue of the snail Lymnaea stagnalis of incorporation GlcNAc by beta 2GlcNAc-transferase I is an essential prerequisite for the action of beta 2GlcNAc-transferase II and beta 2Xyl-transferase. *Eur J Biochem* **232:** 272-283

Mulder H, Schachter H, De Jong-Brink M, Van der Ven JG, Kamerling JP, Vliegenthart JF (1991) Identification of a novel UDP-Gal:GalNAc beta 1-4GlcNAc-R beta 1-3-galactosyltransferase in the connective tissue of the snail Lymnaea stagnalis. *Eur J Biochem* **201:** 459-465

Mulder H, Schachter H, Thomas JR, Halkes KM, Kamerling JP, Vliegenthart JF (1996) Identification of a GDP-Fuc:Gal beta 1-3GalNAc-R (Fuc to Gal) alpha 1-2 fucosyltransferase and a GDP-Fuc:Gal beta 1-4GlcNAc (Fuc to GlcNAc) alpha 1-3 fucosyltransferase in connective tissue of the snail Lymnaea stagnalis. *Glycoconj J* **13:** 107-113

Mulder H, Spronk BA, Schachter H, Neeleman AP, van den Eijnden DH, De Jong-Brink M, Kamerling JP, Vliegenthart JF (1995b) Identification of a novel UDP-GalNAc:GlcNAc beta-R beta 1-4 N-acetylgalactosaminyltransferase from the albumen gland and connective tissue of the snail Lymnaea stagnalis. *Eur J Biochem* **227:** 175-185

Müller R, Hülsmeier AJ, Altmann F, Ten Hagen K, Tiemeyer M, Hennet T (2005) Characterization of mucin-type core-1 beta1-3 galactosyltransferase homologous enzymes in Drosophila melanogaster. *FEBS J* **272:** 4295-4305

Ofek I, Sharon N (1988) Lectinophagocytosis: a molecular mechanism of recognition between cell surface sugars and lectins in the phagocytosis of bacteria. *Infect Immun* **56:** 539-547

Oka T, Jigami Y (2006) Reconstruction of de novo pathway for synthesis of UDP-glucuronic acid and UDP-xylose from intrinsic UDP-glucose in Saccharomyces cerevisiae. *FEBS J* **273:** 2645-2657

Palmberger D, Rendić D, Tauber P, Krammer F, Wilson IB, Grabherr R (2011) Insect cells for antibody production: evaluation of an efficient alternative. *J Biotechnol* **153:** 160-166

Palmberger D, Wilson IB, Berger I, Grabherr R, Rendic D (2012) SweetBac: a new approach for the production of mammalianised glycoproteins in insect cells. *PLoS One* **7:** e34226

Patchett RA, Kelly AF, Kroll RG (1991) The adsorption of bacteria to immobilized lectins. *J Appl Bacteriol* **71:** 277-284

Podolsky DK (1985) Oligosaccharide structures of isolated human colonic mucin species. *J Biol Chem* **260:** 15510-15515

Prokop O, Uhlenbruck G, Köhler W (1968) A new source of antibody-like substances having anti-blood group specificity. A discussion on the specificity of Helix agglutinins. *Vox Sang* **14:** 321-333

Rademacher TW, Parekh RB, Dwek RA, Isenberg D, Rook G, Axford JS, Roitt I (1988) The role of IgG glycoforms in the pathogenesis of rheumatoid arthritis. *Springer Semin Immunopathol* **10:** 231-249

Raja SB, Murali MR, Kumar NK, Devaraj SN (2011) Isolation and partial characterisation of a novel lectin from Aegle marmelos fruit and its effect on adherence and invasion of Shigellae to HT29 cells. *PLoS One* **6:** e16231

Raman J, Guan Y, Perrine CL, Gerken TA, Tabak LA (2012) UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyltransferases: completion of the family tree. *Glycobiology* **22:** 768-777

Rearick JI, Fujimoto K, Kornfeld S (1981) Identification of the mannosyl donors involved in the synthesis of lipid-linked oligosaccharides. *J Biol Chem* **256:** 3762-3769

Sanchez JF, Lescar J, Chazalet V, Audfray A, Gagnon J, Alvarez R, Breton C, Imberty A, Mitchell EP (2006) Biochemical and structural analysis of Helix pomatia agglutinin. A hexameric lectin with a novel fold. *J Biol Chem* **281:** 20171-20180

Schumacher U, Adam E (1997) Lectin histochemical HPA-binding pattern of human breast and colon cancers is associated with metastases formation in severe combined immunodeficient mice. *Histochem J* **29:** 677-684

Schumacher U, Brooks SA, Mester J (2005) The lectin Helix pomatia agglutinin as a marker of metastases--clinical and experimental studies. *Anticancer Res* **25:** 1829-1830

Sharon N, Lis H (1993) Carbohydrates in cell recognition. *Sci Am* **268:** 82-89

Sheriff S, Chang CY, Ezekowitz RA (1994) Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil. *Nat Struct Biol* **1:** 789-794

Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo DH, Larsson T, Lv J, Arendt D, Savage R, Osoegawa K, de Jong P, Grimwood J, Chapman JA,
Spiro RG (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12:** 43R-56R

Sponholtz GM, Short RB (1976) Schistosoma mansoni miracidia: stimulation by calcium and magnesium. *J Parasitol* **62:** 155-157

Stepan H, Bleckmann C, Geyer H, Geyer R, Staudacher E (2010) Determination of 3-O- and 4-O-methylated monosaccharide constituents in snail glycans. *Carbohydr Res* **345:** 1504-1507

Stepan H, Pabst M, Altmann F, Geyer H, Geyer R, Staudacher E (2012) O-Glycosylation of snails. *Glycoconj J* **29:** 189-198

Stepan H, Staudacher E (2011) Optimization of monosaccharide determination using anthranilic acid and 1-phenyl-3-methyl-5-pyrazolone for gastropod analysis. *Anal Biochem* **418:** 24-29

Stibbs HH, Chernin E, Ward S, Karnovsky ML (1976) Magnesium emitted by snails alters swimming behaviour of Schistosoma mansoni miracidia. *Nature* **260:** 702-703

Stwora-Wojczyk MM, Kissinger JC, Spitalnik SL, Wojczyk BS (2004) O-glycosylation in Toxoplasma gondii: identification and analysis of a family of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Int J Parasitol* **34:** 309-322

Summers MD, Smith GE (1987) A manual of methods for baculovirus vectors and insect cell culture procedures. *Texas Agricultural Experiment Station Bulletin*. 1555

Sørensen T, White T, Wandall HH, Kristensen AK, Roepstorff P, Clausen H (1995) UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. Identification and separation of two distinct transferase activities. *J Biol Chem* **270:** 24166-24173

Taus C, Lucini C, Sato T, Furukawa K, Grabherr R, Staudacher E (2013) Expression and characterization of the first snail-derived UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. *Glycoconj J*

Ten Hagen KG, Hagen FK, Balys MM, Beres TM, Van Wuyckhuyse B, Tabak LA (1998) Cloning and expression of a novel, tissue specifically expressed member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family. *J Biol Chem* **273:** 27749-27754

Ten Hagen KG, Tetaert D, Hagen FK, Richet C, Beres TM, Gagnon J, Balys MM, VanWuyckhuyse B, Bedi GS, Degand P, Tabak LA (1999) Characterization of a UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase that displays glycopeptide N-acetylgalactosaminyltransferase activity. *J Biol Chem* **274:** 27867-27874

Ten Hagen KG, Tran DT (2002) A UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase is essential for viability in Drosophila melanogaster. *J Biol Chem* **277:** 22616-22622

Ten Hagen KG, Tran DT, Gerken TA, Stein DS, Zhang Z (2003) Functional characterization and expression analysis of members of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family from Drosophila melanogaster. *J Biol Chem* **278:** 35039-35048

Tenno M, Toba S, Kézdy FJ, Elhammer AP, Kurosaka A (2002) Identification of two cysteine residues involved in the binding of UDP-GalNAc to UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase 1 (GalNAc-T1). *Eur J Biochem* **269:** 4308-4316

Tessier DC, Thomas DY, Khouri HE, Laliberté F, Vernet T (1991) Enhanced secretion from insect cells of a foreign protein fused to the honeybee melittin signal peptide. *Gene* **98:**177-83

Theodoropoulos G, Hicks SJ, Corfield AP, Miller BG, Carrington SD (2001) The role of mucins in host-parasite interactions: Part II - helminth parasites. *Trends Parasitol* **17:** 130-135

van Die I, Cummings RD, van Tetering A, Hokke CH, Koeleman CA, van den Eijnden DH (2000) Identification of a novel UDP-Glc:GlcNAc beta1-->4-glucosyltransferase in Lymnaea stagnalis that may be involved in the synthesis of complex-type oligosaccharide chains. *Glycobiology* **10:** 263-271

van Halbeek H, Strang AM, Lhermitte M, Rahmoune H, Lamblin G, Roussel P (1994) Structures of monosialyl oligosaccharides isolated from the respiratory mucins of a non-secretor (O, Lea+b-) patient suffering from chronic bronchitis. Characterization of a novel type of mucin carbohydrate core structure. *Glycobiology* **4:** 203-219

van Ree R, Cabanes-Macheteau M, Akkerdaas J, Milazzo JP, Loutelier-Bourhis C, Rayon C, Villalba M, Koppelman S, Aalberse R, Rodriguez R, Faye L, Lerouge P (2000) Beta(1,2)-xylose and alpha(1,3)-fucose residues have a strong contribution in IgE binding to plant glycoallergens. *J Biol Chem* **275:** 11451-11458

van Tetering A, Schiphorst WE, van den Eijnden DH, van Die I (1999) Characterization of a core alpha1-->3-fucosyltransferase from the snail Lymnaea stagnalis that is involved in the synthesis of complex-type N-glycans. *FEBS Lett* **461:** 311-314

Vaughn JL, Goodwin RH, Tompkins GJ, McCawley P (1977) The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). *In Vitro* **13:** 213-217

Wandall HH, Hassan H, Mirgorodskaya E, Kristensen AK, Roepstorff P, Bennett EP, Nielsen PA, Hollingsworth MA, Burchell J, Taylor-Papadimitriou J, Clausen H (1997) Substrate specificities of three members of the human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3. *J Biol Chem* **272:** 23503-23514

Wang H, Tachibana K, Zhang Y, Iwasaki H, Kameyama A, Cheng L, Guo J, Hiruma T, Togayachi A, Kudo T, Kikuchi N, Narimatsu H (2003) Cloning and characterization of a novel UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase, pp-GalNAc-T14. *Biochem Biophys Res Commun* **300:** 738-744

Wassarman PM (1990) Profile of a mammalian sperm receptor. *Development* **108:** 1-17

Waterston R, Martin C, Craxton M, Huynh C, Coulson A, Hillier L, Durbin R, Green P, Shownkeen R, Halloran N (1992) A survey of expressed genes in Caenorhabditis elegans. *Nat Genet* **1:** 114-123

Weis WI, Drickamer K (1994) Trimeric structure of a C-type mannose-binding protein. *Structure* **2:** 1227-1240

Weis WI, Drickamer K (1996) Structural basis of lectin-carbohydrate recognition. *Annu Rev Biochem* **65:** 441-473

Weis WI, Taylor ME, Drickamer K (1998) The C-type lectin superfamily in the immune system. *Immunol Rev* **163:** 19-34

Wickham TJ, Nemerow GR (1993) Optimization of growth methods and recombinant protein production in BTI-Tn-5B1-4 insect cells using the baculovirus expression system. *Biotechnol Prog* **9:** 25-30

Wilson IB, Altmann F (1998) Structural analysis of N-glycans from allergenic grass, ragweed and tree pollens: core alpha1,3-linked fucose and xylose present in all pollens examined. *Glycoconj J* **15:** 1055-1070

Wilson IB, Paschinger K, Rendić D (2009) Glycosylation of modeland“Lower”organisms. In: Gabius, H.-J. (ed.) The Sugar Code,pp. 139–154. Wiley-VCH, Weinheim

Wilson R, Ainscough R, Anderson K, Baynes C, Berks M, Bonfield J, Burton J, Connell M, Copsey T, Cooper J (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans. *Nature* **368:** 32-38

Winans KA, Bertozzi CR (2002) An inhibitor of the human UDP-GlcNAc 4-epimerase identified from a uridine-based library: a strategy to inhibit O-linked glycosylation. *Chem Biol* **9:** 113-129

Wopereis S, Lefeber DJ, Morava E, Wevers RA (2006) Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. *Clin Chem* **52:** 574-600

Yuasa HJ, Furuta E, Nakamura A, Takagi T (1998) Cloning and sequencing of three C-type lectins from body surface mucus of the land slug, Incilaria fruhstorferi. *Comp Biochem Physiol B Biochem Mol Biol* **119:** 479-84

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PW, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng

C, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A, Hedgecock D, Xu Z, Liu Y, Domazet-Lošo T, Du Y, Sun X, Zhang S, Liu B, Cheng P, Jiang X, Li J, Fan D, Wang W, Fu W, Wang T, Wang B, Zhang J, Peng Z, Li Y, Li N, Chen M, He Y, Tan F, Song X, Zheng Q, Huang R, Yang H, Du X, Chen L, Yang M, Gaffney PM, Wang S, Luo L, She Z, Ming Y, Huang W, Huang B, Zhang Y, Qu T, Ni P, Miao G, Wang Q, Steinberg CE, Wang H, Qian L, Liu X, Yin Y (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490:** 49-54

Glycoconj J
DOI 10.1007/s10719-013-9486-6

Expression and characterization of the first snail-derived UDP-*N*-acetyl-α-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase

Christopher Taus · Chantal Lucini · Takeshi Sato · Kiyoshi Furukawa · Reingard Grabherr · Erika Staudacher

Received: 11 April 2013 / Revised: 10 June 2013 / Accepted: 3 July 2013

Abstract UDP-GalNAc:polypeptide GalNAc transferase (ppGalNAcT; EC 2.4.1.41) catalyzes the first step in mucin-type O-glycosylation. To date, several members of this large enzyme family have been analyzed in detail. In this study we present cloning, expression and characterization of the first representative of this type of glycosyltransferase from mollusk origin, namely from *Biomphalaria glabrata*. The full length sequence of the respective gene was obtained by screening of a cDNA library using homology-based PCR. The entire gene codes for a protein consisting of 600 amino acids comprising the features of a typical type II membrane protein containing a cytoplasmic tail at the N-terminus, a transmembrane and a catalytic domain as well as a ricin-like motif at the C-terminus. Sequence comparison with ppGalNAcTs from various species revealed high similarities in terms of structural architecture. The enzyme is O-glycosylated but does not have any putative N-glycosylation sites. All four tested acceptor peptides were functional substrates, with Muc2 being the best one. Further biochemical parameters tested, confirmed a close relationship to the family of yet known ppGalNAcTs.

Keywords ppGalNAcT · GalNAc-transferase · O-glycosylation · Snail · O-glycoprotein · Biomphalaria glabrata

Abbreviations

Bge cells	*Biomphalaria glabrata* embryonic cells
GalNAc	*N*-acetylgalactosamine
GlcNAc	*N*-acetylglucosamine
MES	2-(*N*-morpholino)ethanesulphonic acid
ppGalNAcT	polypeptide *N*-acetylgalactosaminyltransferase
Sf9-cells	*Spodoptera frugiperda* cells

Introduction

Mucin-type O-glycosylation is the most common type of O-glycosylation. Besides mammals it has been found in insects, snails, worms, and various parasites [1, 2]. The initiating enzymes of this type of glycosylation are members of an evolutionarily conserved family of UDP-GalNAc:polypeptide GalNAc transferases (ppGalNAcTs, [EC 2.4.1.41]), which transfer a GalNAc residue from UDP-GalNAc to a serine (Ser) or threonine (Thr) within a polypeptide chain. Depending on the organism, tissue and developmental stage, the protein bound GalNAc residue is elongated first by a GlcNAc and/or galactose residue and subsequently is modified by more of these or other monosaccharides (fucose or sialic acid). Mucin-type O-glycosylation has been suggested to influence conserved processes in development and changes in O-glycosylation patterns have been connected with pathogenic

C. Taus · C. Lucini · E. Staudacher
Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Muthgasse 18, 1190 Vienna, Austria

T. Sato · K. Furukawa
Laboratory of Glycobiology, Department of Bioengineering, Nagaoka University of Technology, Kamitomioka 1603-1, Niigata 940-2188, Nagaoka, Japan

R. Grabherr
Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Muthgasse 18, 1190 Vienna, Austria

E. Staudacher (✉)
Department of Chemistry, Glycobiology, University of Natural Resources and Life Sciences, Vienna, Muthgasse 18, 1190 Vienna, Austria
e-mail: erika.staudacher@boku.ac.at

Published online: 23 July 2013

events. Therefore, understanding the biosynthesis of these structures, as well as their biological role is of high interest in basic and applied research.

In humans already 20 representatives of this enzyme family have been identified and several more in other organisms. They have been characterized and grouped into subfamilies according to conserved structural as well as functional aspects [3]. These sub-families differ in details of structure, tissue distribution and their acceptor peptide preferences. Yet, some common structural elements have been identified: ppGalNAcT is a type II membrane bound glycosyltransferase located in the Golgi apparatus with a highly conserved UDP-GalNAc-binding region, a manganese-binding site and quite often a lectin-like domain located at the carboxyl terminus, which seems to have a major influence on substrate specificity [4, 5].

So far, only few invertebrates have been characterized in terms of their ppGalNAcT activities. As a representative of insects *Drosophila melanogaster* has been found to contain 14 members of this enzyme family with some of them being essential for normal development [6, 7]. In the nematode *Caenorhabditis elegans* 11 homologues of vertebrate ppGalNAcTs were identified showing 60–80 % amino acid sequence similarity [8]. Also parasites (*Fasciola hepatica*, *Trypanosoma cruzi* and *Toxoplasma gondii*) contain ppGalNAcTs initiating their O-glycosylation [9–11]. The recent review by Bennett [3] gives a comprehensive coverage of the ppGalNAcTs that have been discovered so far in different organisms.

In the course of our previous work, we identified small mucin type O-glycans containing a protein bound GalNAc-residue elongated by galactose residues and, in some cases also fucose residues, in several snail species (*Arion lusitanicus*, *Achatina fulica*, *Biomphalaria glabrata*, *Cepaea hortensis*, *Clea helena*, *Helix pomatia*, *Limax maximus* and *Planorbarius corneus*) [2]. Many of the terminal hexoses were found to be methylated which is a typical structural element in snails [12, 13].

In the current study we present the identification, molecular cloning and characterization of the first ppGalNAcT from mollusk origin, in particular from *Biomphalaria glabrata*, a snail living in fresh water. The identified enzyme initiates mucin-type O-glycosylation and shares all important domains with the other members of the family. Our findings are another proof that O-glycosylation is a highly conserved process of protein modification throughout the animal kingdom.

Material and methods

Materials All DNA manipulations were carried out essentially as summarized by Sambrook *et al*. [14], restriction enzymes, T4 ligase, T4 polynucleotide kinase, DNA Polymerase I, Large (Klenow) Fragment and calf intestine alkaline phosphatase were purchased from New England Biolabs (Frankfurt, Germany). All enzymes were used according to manufacturers' recommendations. All primers were synthesized commercially (Sigma-Aldrich, Vienna, Austria). Embryonic cells from *Biomphalaria glabrata* (Bge cells, NR-40248, BEI Resources, NIAID, NIH) were grown as described at 26 °C [15]. Chemically transformed *Escherichia coli* (High Efficiency) cells (New England Biolabs, Frankfurt, Germany), were spread on Lysogeny broth (LB) agar plates containing 100 μg/ml Ampicillin and incubated overnight at 37 °C. *Spodoptera frugiperda* cells (Sf9, ATCC CRL-1711) were cultivated in IPL41 medium (SAFC Biosciences, St. Louis, USA) containing yeast extract, a lipid mixture supplemented with 10 % fetal calf serum, at 27 °C [16].

Synthesis of a cDNA library from Biomphalaria glabrata Total RNA was isolated from Bge cells by RNeasy Mini Kit (Qiagen, Hilden, Germany). Polyadenylated RNA was purified from 12 μg of total RNA by MicroPoly(A)Purist^{TM} Kit (Ambion, Austin, USA). 100 ng of polyadenylated RNA were used as a template for first strand cDNA synthesis by In-Fusion® SMARTer™ Directional cDNA Library Construction kit (Takara, Saint-Germain-en-Laye, France). Double-stranded cDNA was synthesized according to the manufacturer's manual, ligated and transformed chemically into NEB 10-beta Competent *Escherichia coli* cells. Recombinant *E. coli* colonies were used as a template for control PCR using the primer set 5′TCACACAGGAAACAGCTATGA3′ and 5′CCTCTTCGCTATTACGCCAGC3′.

Homology-based PCR Highly conserved regions were identified by alignment of ppGalNAcT genes from *Homo sapiens* 2 (NP_004472), *Mus musculus* (NP_038842), *Caenorhabditis elegans* GLY-3 (NP_498722) and *Drosophila melanogaster* (AAQ56699). Based on the conserved amino acid sequences WGGEN and VWMDEY/F sense and antisense degenerated primers (5′TGGGGWGGWGARAACYT3′ and 5′TGWAYTCRTCCATCCADAC3′) were designed. The resulting 187 base pair product was purified by NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel, Düren, Germany), ligated with the pGEM-T Easy vector (Promega, Fitchburg, USA) and sequenced to examine identity of the fragment.

Isolation of the full-length snail ppGalNAcT gene Recovery of the complete sequence of this novel ppGalNAcT gene was obtained by inverse PCR of self-ligated cDNA, as well as a 5′ and 3′ RACE PCR, respectively. Double-stranded cDNA (3.5 μg) was precipitated with ethanol, resuspended in

NEBuffer 4 and incubated with DNA Polymerase I, Large (Klenow) Fragment in a total volume of 25 μl for 15 min at 25 °C. After DNA purification and ethanol precipitation the cDNA was treated with 10 units of T4 Polynucleotide Kinase in a total volume of 30 μl for 30 min at 37 °C. The enzyme was inactivated at 65 °C for 20 min. T4 DNA Ligase was used for the self-ligation reaction (16 °C, over night). Correct self-ligated cDNAs coding for the ppGalNAcT gene were amplified by inverse PCR using the primer set 5′GAA ACACAAAGCGTGCAGCAGAAG3′ and 5′CACTGCCA AACACGGAAGGATATC3′. The amplified products were purified and ligated with the pGEM-T Easy vector for sequencing.

5′ and 3′ Ready RACE cDNA was synthesized from 1 μg of total RNA by using the SMARTer™ RACE cDNA Amplification Kit (Takara, Saint-Germain-en-Laye, France). The complete 5′ end of the novel ppGalNAcT gene was amplified by using the Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Vienna, Austria), 5′ Ready RACE cDNA as template, the gene specific primer 5′CATCCAGACTTCTG CTGCACGCTTTGTG3′ and the 10X Universal Primer A Mix provided by the kit. The amplified 5′ end was purified and ligated with pUC19 vector for sequencing. A nested PCR resulted in the amplification of the complete 3′ end using ExTaq Polymerase (Takara, Saint-Germain-en-Laye, France) and 3′ Ready RACE cDNA as template. The amplified 3′ end was purified and ligated with pGEM-T Easy vector for sequencing.

The nucleotide sequence reported in this paper has been deposited in the GenBank database: *GenBank*: *KC18251*.

Recombinant expression, purification and Western blot analysis A cDNA fragment without cytoplasmic tail and transmembrane (amino acid 1–26) domain was amplified by using the forward primer 5′GGGGAGCTCGGAGATGAT CAAAGTGAGTTTG3′ and the reverse primer 5′GGGGGT ACCCTATCTGTTTTTACTTAAAGAAAATGTCCAC3′. The purified ppGalNAcT fragment was ligated with the pGEM-T Easy vector, cut, purified and ligated to pVT-Bac [17]. To verify the gene sequence and the correct insertion into the cloning site, this construct was sequenced. 2 μg of the recombinant construct (containing the honeybee melittin secretion signal, a 6 x His tag and the ppGalNAcT fragment) were co-transfected with 300 ng linearized BD BaculoGold™ Bright Baculovirus DNA (Becton Dickinson, Schwechat, Austria) into Sf9 insect cells. The secreted recombinant protein was purified on Ni-NTA agarose (Qiagen, Hilden, Germany) and analyzed by SDS-PAGE and Western blotting using Penta-His monoclonal antibody (Qiagen, Hilden, Germany; dilution 1:2000) followed by alkaline phosphatase conjugated anti-mouse IgG from goat (Sigma-Aldrich, Vienna, Austria; dilution 1:3000) [18]. Fetuin was used as a negative control for Penta-His monoclonal antibody.

Analysis of the recombinant ppGalNAcT activity The enzyme activity of the recombinant ppGalNAcT was determined in 20 μl reaction mixture containing 50 mM MES (2-(*N*-morpholino)ethanesulfonic acid), pH 7.0, 10 mM $MnCl_2$, 0.1 % Triton X-100, 40 nmol UDP-GalNAc (Sigma-Aldrich, Vienna, Austria), 20 μg acceptor peptide (Cellmano Biotech Co., Ltd., Shanghai, China) and 2 μl enzyme solution (protein concentration 5.75 μg/ml) at 37 °C for 45 min. The peptides were chosen randomly: Muc1a (APPAHGVTSAPDTRPA PGC) and 1a' (AHGVTSAPDTR) were taken from [19], Muc2 (PTTTPITTTTTVTPTPTPTGTQTK) from [20] and Muc5Ac (GTTPSPVPTTSTTSAP) from [21]. The reaction was terminated by boiling at 96 °C for 5 min and analyzed by HPLC on a reversed phase C18 column (4.6×250 mm, 5 μm, Thermo Scientific, Vienna, Austria) in 0.1 % trifluoroacetic acid in water, applying a linear gradient from 10 to 30 % of eluent (0,1 % trifluoroacetic acid in acetonitrile) in 25 min at a flow rate of 1 ml/min. The long term incubation was carried out for 24 h analogous to the standard assay with additions of 10 nmol UDP-GalNAc and 2 ul of enzyme solution every 3 hours.

For the analysis of the biochemical parameters the standard assay conditions using Muc2 as acceptor peptide were modified as follows. For determination of the manganese optimum the concentration of $MnCl_2$ was varied from 0–300 mM, for the determination of cation requirement the standard assay was carried out without any cation addition or in the presence of 10 mM of EDTA, Mn^{2+}, Mg^{2+}, Ca^{2+}, Co^{2+}, Cu^{2+}, Ni^{2+}, or Ba^{2+}. The determination of enzyme stability and the pH-optimum were done according to Peyer *et al.* [22]. Kinetic data were acquired using 1:4 diluted enzyme solution and Muc2 at eleven different concentrations ranging from 0.01 to 0.82 mM, with UDP-GalNAc kept constant at a concentration of 2 mM. Similarly UDP-GalNAc was varied between concentrations from 0.01 to 2 mM while Muc2 was kept constant at 0.82 mM. Km-values were obtained from Lineweaver-Burk plots.

All quantitative values were calculated from the area of HPLC patterns. Each assay was carried out at least in duplicate with appropriate controls.

Determination of protein content Protein concentrations were determined by the Micro-BCA protein assay (Pierce, Bonn, Germany) with bovine serum albumin as the standard.

MALDI-TOF MS analysis MALDI-TOF MS analysis was carried out on an Autoflex Speed MALDI-TOF (Bruker Daltonics, Germany) equipped with a 1,000 Hz Smartbeam.II laser in positive mode using α-cyano-4-hydroxycinnamic acid

as matrix. Spectra were processed with the manufacturer's software (Bruker Flexanalysis 3.3.80).

Results

Isolation of the full-length ppGalNAcT gene A full-length cDNA library from Bge cells was created in order to isolate and characterize the first mollusk ppGalNAcT. The double-stranded cDNA synthesized from purified polyadenylated RNA was cloned and propagated in *E.coli*, resulting in a full length cDNA library. A control PCR of 50 randomly selected colonies showed high diversity with inserts ranging from about 500 up to 9,000 base pairs (data not shown).

Our PCR strategy was based on homology to the conserved regions (amino acid sequence WGGEN and VWMDEY/F) of four different ppGalNAcTs (*Homo sapiens* 2 NP_004472, *Mus musculus* NP_038842, *Caenorhabditis elegans* NP_498722 and *Drosophila melanogaster* AAQ56699). The worm and the insect enzyme were chosen due to their estimated closeness to mollusks, the two mammalian enzymes were chosen arbitrarily. By using degenerated primers an 187 base pair product was amplified. In order to receive the full-length ppGalNAcT gene, an inverse PCR of self-ligated cDNA was performed, which resulted in the amplification of a 1,659 base pair product coding for 553 amino acids (6 – 558). The remaining 5′ (amino acid 0 – 417) and 3′ end (amino acid 481 – 600) was amplified by RACE-PCR. Thereby, the full-length ppGalNAcT gene could be recovered, which turned out to be coding for a 600 amino acid type-II membrane protein containing a putative N-terminal cytoplasmic tail, a transmembrane domain (amino acid 7–25, predicted by TMHMM Server v. 2.0), a stem region, a luminal catalytic domain and a ricin-like motif at the C-terminus similar to almost all of the ppGalNAcTs described.

Purification and determination of specificity A truncated version of the Bge ppGalNAcT, without transmembrane domain and with the addition of an insect specific secretion signal, was expressed in Sf9 cells. The supernatant was tested for ppGalNAcT enzyme activity, as well as the supernatant of Sf9 cells expressing another glycosyltransferase. Whereas the supernatant of cells infected with ppGalNAcT expressing baculoviruses showed high enzyme activity, no transfer at all was detected in the control supernatant, confirming no interfering ppGalNAcT activity derived from the baculovirus insect cell expression system. Recombinant ppGalNAcT was further purified using metal chelate affinity chromatography. Quality of the purification was determined by Coomassie staining and immunoblotting using anti-His antibody. A band at approximately 65 kDa was detected, which correlated with the molecular weight as calculated by the sum of the amino acids of the truncated form (Figs. 1a and b). For further characterization, the purified sample (specific activity 46 milliunits/ml; one unit is defined as the amount of enzyme that transfers 1 μmol of GalNAc in 1 min of the standard reaction mixture) was subjected to enzyme activity assays. Four different peptides, which had been successfully used in previous studies, were chosen as acceptors for enzyme activity. Qualitative analysis was done by MALDI-TOF MS; for quantification of the substrate conversion the relative amount of peak areas of the HPLC patterns were used for calculation.

All four arbitrarily selected peptides were functional acceptor substrates indicating that the snail derived ppGalNAcT displays a broad specificity. Similar to already described ppGalNAcTs, we observed the transfer of more than one GalNAc residue upon extended incubation times. Muc2 turned out to be the best acceptor where the addition of a second GalNAc residue was detectable already after 90 min and clearly visible after 120 min (Fig. 2a–d and 3). Using higher enzyme concentrations extension by a third GalNAc residue could be visualized within the standard incubation time of 2 h. By increasing the incubation time up to 24 h it was possible to detect a transfer of 8 GalNAc residues onto Muc2 acceptor peptide (Fig. 2e–h). On Muc5Ac, which was the second best acceptor peptide, traces of a second sugar residue could be seen after 2 h (Fig. 4), but less than 50 % of Muc1a or Muc1a' were converted into monoglycosylated peptides within 2 h of incubation time.

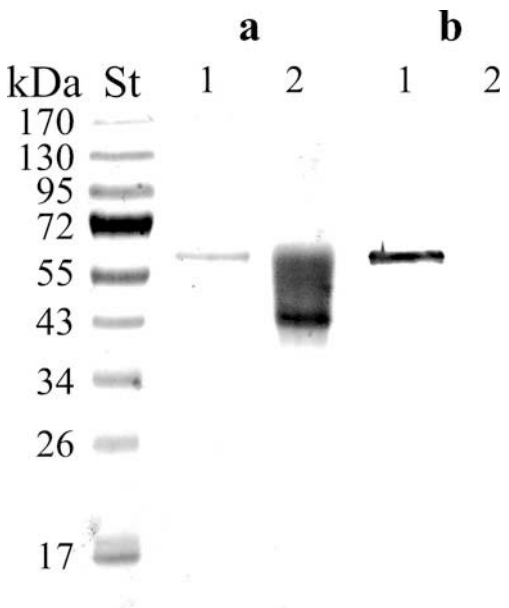

Fig. 1 Purification control of recombinant snail ppGalNAcT by SDS-PAGE and Western blot analysis (lane 1), **a** Coomassie staining **b** Immunoblotting of purified ppGalNAcT with Penta-His monoclonal antibody; Fetuin (lane 2) was used as a negative control for Penta-His monoclonal antibody

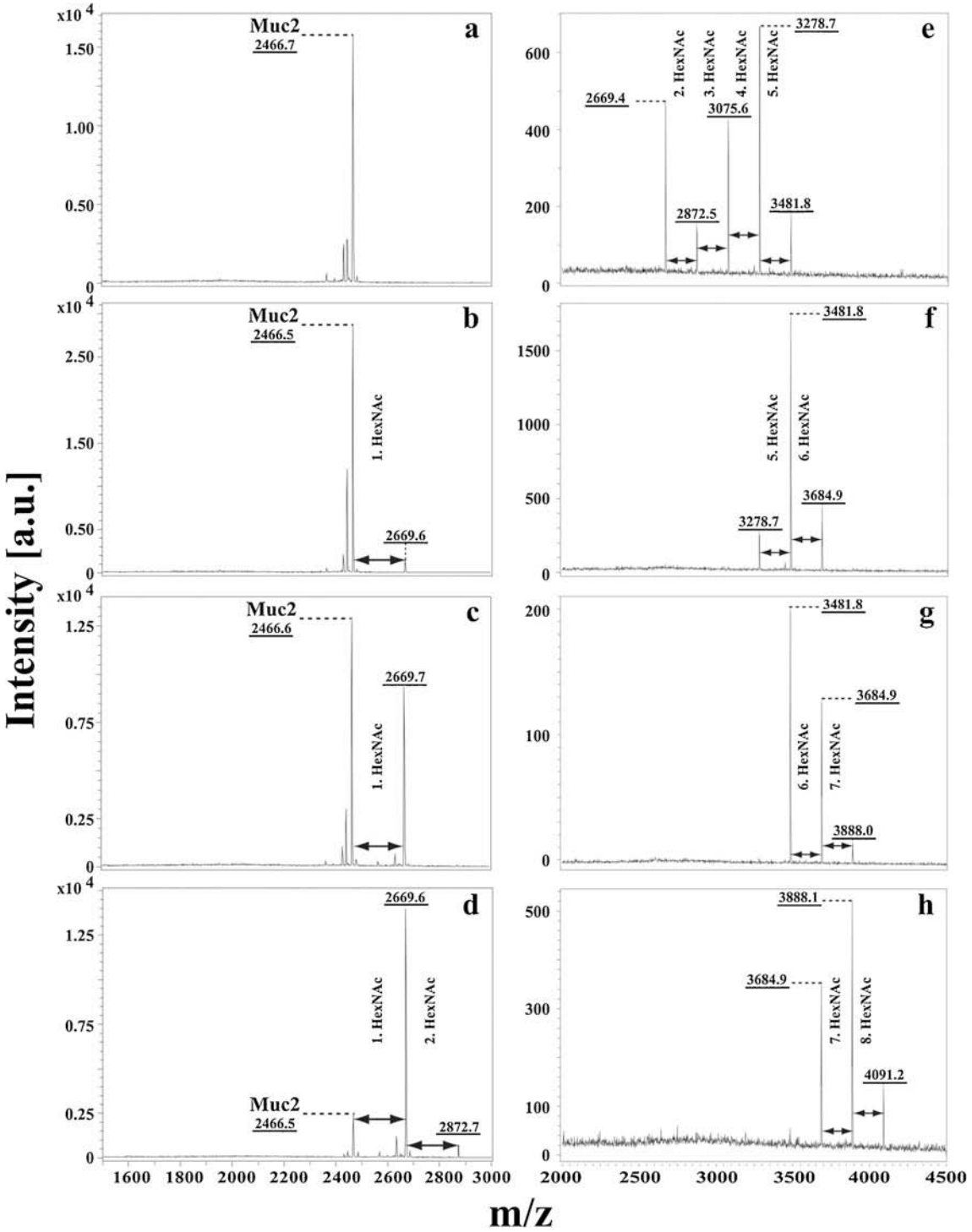

Fig. 2 Mass spectrometric ($[M+Na]^+$) analysis of the transfer of GalNAc using Muc2 as acceptor peptide. Muc2 peptide (**a**); Assays with incubation times of 10 min (**b**), 60 min (**c**), 2 h (**d**), 3 h (**e**), 6 h (**f**), 9 h (**g**) and 24 h (**h**)

Properties of the enzyme According to the amino acid sequence no N-glycosylation site was expected, however, O-glycans linked to the recombinant enzyme expressed in insect cells could be detected by lectin blot analysis using peanut agglutinin in combination with O-glycosidase treatment (data not shown).

The activity of the recombinant ppGalNAcT was not affected by storage for 72 h in a temperature range from −20 °C to room temperature, or by the addition of up to 10 % of methanol or glycerol. Addition of 5 % of acetonitrile had no negative effect on enzyme activity, whereas 10 % of acetonitrile or lyophilization reduced enzyme activity to

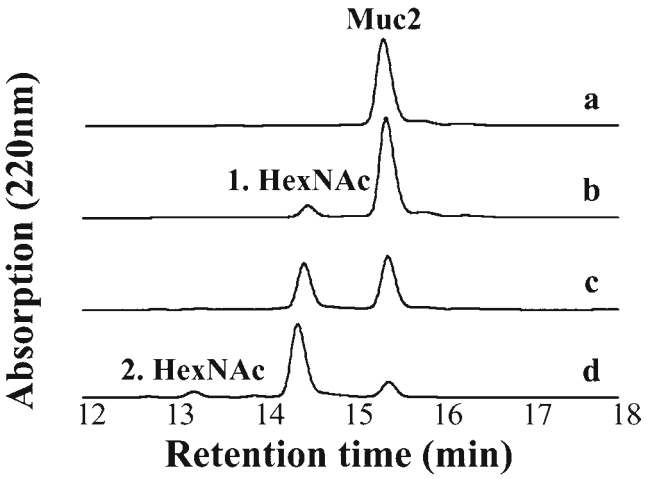

Fig. 3 HPLC analysis of the transfer of GalNAc onto Muc2 acceptor peptide. Muc2 peptide (**a**), incubation for 10 min (**b**), 60 min (**c**) and 2 h (**d**)

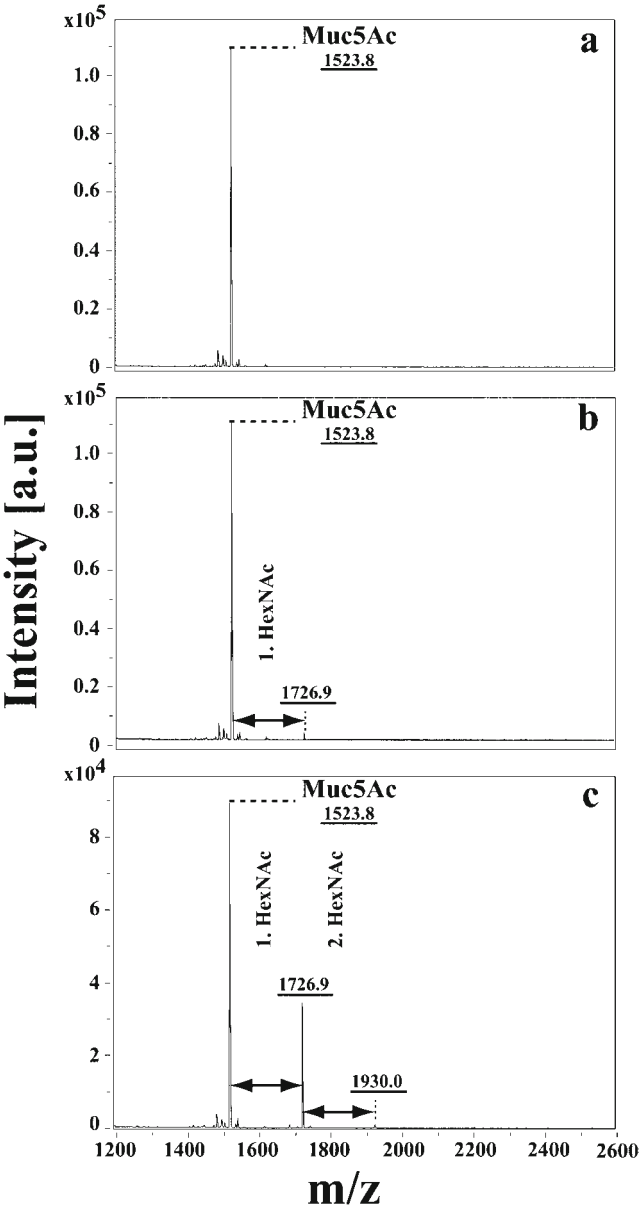

Fig. 4 Mass spectrometric ($[M+Na]^+$) analysis of the transfer of GalNAc onto Muc5Ac acceptor peptide. Muc5Ac peptide (**a**); Standard assay conditions using Muc5Ac as acceptor with incubation times of 20 min (**b**) and 120 min (**c**)

about 60 %. The activity was drastically reduced at storage temperatures above room temperature. After 72 h at 37 °C less than 10 % of activity was detectable. However, 37 °C was the optimal incubation temperature for short assays up to 2 h (data not shown). Further, ppGalNAcT showed a pH-optimum at 6.0–6.5 using MES as the appropriate buffer salt (Fig. 5a). The enzymatic transfer was dependent on divalent cations (no activity in the presence of EDTA) with the order of increasing support being $Mn^{2+}>Co^{2+}>Mg^{2+}>Ca^{2+}$. Cu^{2+}-ions abolished activity completely (Fig. 5b). Standard assays resulting in nearly 100 % of conversion into product were also carried out with half amount of enzyme to confirm the validity of the original assay. Maximal rates of transfer were achieved with Mn^{2+}-concentrations from 10–20 mM. Concentrations above 80 mM reduced the enzyme activity.

Analysis of enzyme activity revealed a K_M of 0.064 mM, with V_{max} at 0,16 nmol/min/μg for Muc2 acceptor peptide and a K_M of 0.046 mM and V_{max} at 0,054 nmol/min/μg for UDP-GalNAc in the presence of constant amounts of Muc2. Both values are within the range of published data (K_M 0.01–0.23 mM [19, 23]) for human ppGalNAc T2 using Muc2 acceptor, however the experiments have been performend under slightly different conditions. No kinetic data are available for the closely related enzymes from other invertebrates.

Discussion and conclusion

ppGalNAcTs are the largest family of glycosyltransferases, which catalyze a single linkage. They can be found throughout the animal kingdom, *e.g.* in mammals, insects and worms and are highly conserved regarding their amino acid sequence.

Mollusks are a large and diverse phylum of invertebrates, living in marine and freshwater as well as in terrestrial habitats. They span a wide diversity in terms of body size, body shape and the size of the brain. Thus, developmental processes are expected to be regulated quite differently. Yet, findings of conserved pathways sharing a common function with other organisms points towards a common origin in the biological past.

We previously found O-glycosylation with GalNAc being the protein bound sugar in several snail tissues [2] and thus, we expected to be able to identify a ppGalNAcT gene representative of this enzyme family from snail origin. Though, at the time we started our project, no complete mollusk genome was available. There were just parts of *Biomphalaria glabrata*

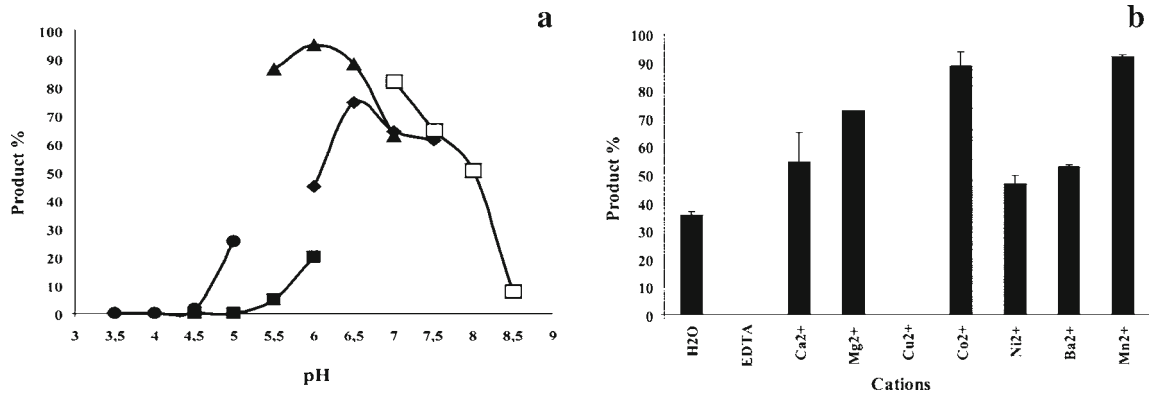

Fig. 5 Properties of ppGalNAcT from Bge cell line. **a** pH pattern using the following buffer systems: (●–●) acetate/NaOH, (■–■) citrate/NaOH, (▲–▲) MES, (♦–♦) phosphate and (□–□) Tris/HCl; **b** Influence of divalent cations on enzyme activity

(*Biomphalaria glabrata* genome initiative; http://biology.unm.edu/biomphalaria-genome/) and *Lottia gigantea* (http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html) sequenced. Data existed from *Lymnea stagnalis*, which described a GalNAcT activity acting on terminal residues of N-glycans [24]. This enzyme had been characterized but not produced recombinantly. In the meantime, the complete sequence from Pacific oyster (*Crassostrea gigas*) has been published but no expression or characterization of glycosyltransferases has been performed [25].

In search for the snail ppGalNAcT gene we started by screening a full-length cDNA library, which we had generated from a *Biomphalaria glabrata* Bge cell line, with degenerated primers that were homologous to highly conserved parts of ppGalNAcTs that were published so far. As concluded from sequence comparisons, the first piece of sequence obtained (coding for the amino acids 359–420) showed high homology to ppGalNAcTs from several other organisms (*e.g.* 92 % identity to *Crassostrea gigas* (EKC38600), 92 % to *Homo sapiens* 2 (NP_004472), 65 % to *Drosophila melanogaster* (AAQ56699) and 76 % to *Caenorhabditis elegans* (NP_498722)). After three further PCR steps we succeeded in the elucidation of the complete sequence of the full length snail derived ppGalNAcT, comprising a total of 600 amino acids. In the first part of the catalytic domain the glycosyltransferase 1 (GT1) structural motif (residues 166–272) described by [26], containing the conserved $D^{252}xH^{254}$ sequence was identified. This conserved motif has been shown to be involved in Mn^{2+} coordination [27]. C-terminal of the $D^{252}xH^{254}$ sequence there were two cysteine residues (255 and 257) located, which have

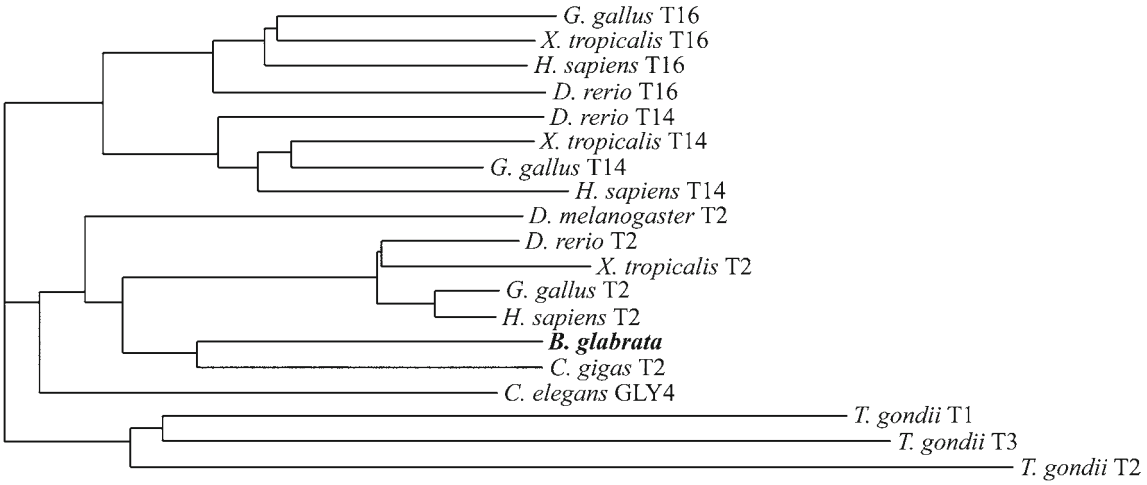

Fig. 6 Branch (group Ib according to [3]) of the phylogenetic tree of ppGalNAcTs. The phylogenetic tree is based on the neighbor-joining method based on amino acid sequences using ClustalW software. The branch length represents evolutionary distance between the members of the family ppGalNAcT enzymes. The following Genbank accession numbers were used: *Gallus gallus* T16 (XP_001231965.1), *Xenopus tropicalis* T16 (NP_001039091.1), *Homo sapiens* T16 (AJ505951), *Danio rerio* T16 (XP_001339749.3), *Danio rerio* T14 (NP_001038460.1), *Xenopus tropicalis* T14 (NP_001072369), *Gallus gallus* T14 (XM_419370.2), *Homo sapiens* T14 (Y09324), *Drosophila melanogaster* T2 (NP_608773.2), *Danio rerio* T2 (NP_001121823.1), *Xenopus tropicalis* T2 (XP_002931524.1), *Gallus gallus* T2 (XP_419581.2), *Homo sapiens* T2 (X85019), *Biomphalaria glabrata* (KC182513), *Crassostrea gigas* T2 (EKC38600), *Caenorhabditis elegans* GLY4 (NP_507850.2), *Toxoplasma gondii* T1 (XP_002365147.1), *Toxoplasma gondii* T3 (XP_002369811.1), *Toxoplasma gondii* T2 (XP_002365091.1)

been reported to be responsible for UDP-binding [28]. The second part of the catalytic domain contained the D^{353}xxxxxWGGENLE^{365} sequence within the Gal/GalNAc-transferase motif, which is shared among ppGalNAcTs and ß1,4-galactosyltransferases [26]. It has been shown that this domain interacts with the GalNAc moiety of the sugar donor [27]. Similar to most of the other ppGalNAcTs the C-terminus of the snail enzyme comprised a lectin domain, which consisted of three homologous repeats (α, ß and γ) belonging to the ricin-type lectin structural family. It has been hypothesized that this domain is involved in recognizing carbohydrate moieties on glycopeptide substrates [26, 29].

An extended phylogenetic tree with 102 ppGalNAcTs from vertebrate and invertebrate sources has been presented recently [3]. There, the new ppGalNAcT from Bge cells fits according to its amino acid homology into the subfamily I, group b (Fig. 6). As expected, the snail enzyme is closely related to the ppGalNAcT from *C. gigas*, another representative from the mollusk kingdom. Furthermore, the snail ppGalNAcT is highly homologous to the T2-enzymes from many other species (human, chicken, frog, fish, fly).

For activity and specificity tests the truncated version of the enzyme without the transmembrane domain was expressed in insect cells. To allow easy purification from the supernatant a honeybee melittin secretion signal and a His tag were fused to the N-terminus. The Bge ppGalNAcT was able to recognize all four chosen peptides as acceptor substrates, Muc2, Muc5Ac, Muc1a and Muc1a'. For Muc2 and Muc5Ac we observed a multiple GalNAc transfer within a reaction time of two hours of the standard assay, whereas Muc1 and Muc1a' were possible, but rather weak acceptors. Increasing the incubation time up to 24 hours in combination with several additions of enzyme yielded in an incorporation of 8 GalNAc-residues into Muc2. These data demonstrate that the enzyme has a broad specificity and is capable of utilizing non-glycosylated as well as mono- or multi-glycosylated substrates. Comparing the specificity of the snail enzyme with human enzymes, especially with human ppGalNAc T2 and ppGalNAc T14 [20], which are most close to our snail enzyme from the structural point of view (Fig. 6), the results indicate that Bge ppGalNAcT is a member of a ppGalNAc T2 subgroup. This is also supported by comparing the specificity with the insect ppGalNAc T2 from *Drosophila* [30]

Based on the amino acid sequence of the enzyme no potential N-glycosylation site was predicted. The recombinant enzyme expressed in Sf9-cells was O-glycosylated (detected by peanut agglutinin staining and O-glycosidase sensitivity; data not shown). We do not know the native glycosylation status of this protein in the snail, but in principle O-glycosylation in snails is possible [2].

In terms of biochemical parameters the Bge ppGalNAcT correlated with those determined for other members of the family. Its stability performance was comparable as well as the essential requirement of divalent cations with Mn^{2+} being optimal in a range of 10–20 mM. However, while most of the other enzymes have a pH-optimum at 7.0–7.5, the Bge ppGalNAcT shows its highest activity at pH 6.0–6.5.

In this study we could identify, clone, express and analyze a ppGalNAcT from gastropod origin. It is the first expressed and fully characterized enzyme of this family from the mollusk kingdom. According to its structure and its substrate specificity it is a close relative to ppGalNAcT2s from various sources.

Acknowledgments This work was supported by the Austrian Science Fund (FWF): project number P22118-B20. The authors want to thank Dr. Regina Grillari and Dr. Dubravko Rendić for their help with the Bge cells.

References

1. Wilson, I.B.H., Paschinger, K., Rendić, D.: Glycosylation of model and "Lower" organisms. In: Gabius, H.-J. (ed.) The Sugar Code, pp. 139–154. Wiley-VCH, Weinheim (2009)
2. Stepan, H., Pabst, M., Altmann, F., Geyer, H., Geyer, R., Staudacher, E.: O-Glycosylation of snails. Glycoconjugate J **29**, 189–198 (2012)
3. Bennett, E.P., Mandel, U., Clausen, H., Gerken, T.A., Fritz, T.A., Tabak, L.A.: Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology **22**, 736–756 (2012)
4. Brockhausen, I., Schachter, H., Stanley, P.: O-GalNAc glycans. In: Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E. (eds.) Essentials of Glycobiology, pp. 115–127. Cold Spring Harbor, Laboratory Press, New York (2008)
5. Raman, J., Guan, Y., Perrine, C.L., Gerken, T.A., Tabak, L.A.: UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyltransferases: completion of the family tree. Glycobiology **22**, 768–777 (2012)
6. Ten Hagen, K.G., Tran, D.T.: A UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase isessential for viability in *Drosophila melanogaster*. J. Biol. Chem. **277**, 22616–22622 (2002)
7. Ten Hagen, K.G., Hagen, F.K., Balys, M.M., Beres, T.M., Van Wuyckhuyse, B., Tabak, L.A.: Cloning and expression of a novel, tissue specifically expressed member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family. J. Biol. Chem. **273**, 27749–27754 (1998)
8. Hagen, F.K., Nehrke, K.: cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase sequence homologs from *Caenorhabditis elegans*. J. Biol. Chem. **273**, 8268–8277 (1998)
9. Freire, T., Casaravilla, C., Carmona, C., Osinaga, E.: Mucin-type O-glycosylation in *Fasciola hepatica*: characterisation of carcinoma-associated Tn and sialyl-Tn antigens and evaluation of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase activity. Int. J. Parasitol. **33**, 47–56 (2003)
10. Freire, T., Robello, C., Soulé, S., Ferreira, R., Osinaga, E.: Sialyl-Tn antigen expression and O-linked GalNAc-Thr synthesis by *Trypanosoma cruzi*. Biochem. Biophys. Res. Commun. **312**, 1309–1316 (2003)

11. Stwora-Wojczyk, M.M., Kissinger, J.C., Spitalnik, S.L., Wojczyk, B.S.: O-glycosylation in *Toxoplasma gondii*: identification and analysis of a family of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. Int. J. Parasitol. **34**, 309–322 (2004)
12. Stepan, H., Bleckmann, C., Geyer, H., Geyer, R., Staudacher, E.: Determination of 3-O- and 4-O-methylated monosaccharide constituents in snail glycans. Carbohydr. Res. **345**, 1504–1507 (2010)
13. Staudacher, E.: Methylation - an uncommon modification of glycans. Biol. Chem. **393**, 675–685 (2012)
14. Sambrook, J., Fritsch, E.F., Maniatis, T.: Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989)
15. Knight, M., Ittiprasert, W., Odoemelam, E.C., Adema, C.M., Miller, A., Raghavan, N., Bridger, J.M.: Non-random organization of the *Biomphalaria glabrata* genome in interphase Bge cells and the spatial repositioning of activated genes in cells co-cultured with *Schistosoma mansoni*. Int. J. Parasitol. **41**, 61–70 (2011)
16. Summers, M.D., Smith, G.E.: A manual of methods for baculovirus vectors and insect cell culture procedures. Tex. Agric. Exp. Station. Bull. 1555 (1987)
17. Tessier, D.C., Thomas, D.Y., Khouri, H.E., Laliberté, F., Vernet, T.: Enhanced secretion from insect cells of a foreign protein fused to the honeybee melittin signal peptide. Gene **98**, 177–183 (1991)
18. Gutternigg, M., Bürgmayr, S., Pöltl, G., Rudolf, J., Staudacher, E.: Neutral N-glycan patterns of the gastropods *Limax maximus*, *Cepaea hortensis*, *Planorbarius corneus*, *Arianta arbustorum* and *Achatina fulica*. Glycoconjugate J. **24**, 475–489 (2007)
19. Wandall, H.H., Mirgorodskaya, E., Kristensen, A.K., Roepstorff, P., Bennett, E.P., Nielsen, P.A., Hollingsworth, M.A., Burchell, J., Taylor-Papadimitriou, J., Clausen, H.: Substrate specificities of three members of the human UDP-N-acetyl-alpha-D-galactosamine: Polypeptide N-acetylgalactosaminyltransferase family. GalNAc-T1, -T2, and -T3. J. Biol. Chem. **272**, 23503–23514 (1997)
20. Wang, H., Tachibana, K., Zhang, Y., Iwasaki, H., Kameyama, A., Cheng, L., Guo, J., Hiruma, T., Togayachi, A., Kudo, T., Kikuchi, N., Narimatsu, H.: Cloning and characterization of a novel UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase pp-GalNAc-T14. Biochem. Biophys. Res. Commun. **300**, 738–744 (2003)
21. Ten Hagen, K.G., Tetaert, D., Hagen, F.K., Richet, C., Beres, T.M., Gagnon, J., Balys, M.M., VanWuyckhuyse, B., Bedi, G.S., Degand, P., Tabak, L.A.: Characterization of a UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase that displays glycopeptide N-acetylgalactosaminyltransferase activity. J. Biol. Chem. **274**, 27867–27874 (1999)
22. Peyer, C., Bonay, P., Staudacher, E.: Purification and characterization of a β-xylosidase from potatoes (*Solanum tuberosum*). Biochim. Biophys. Acta **1672**, 27–35 (2004)
23. Sørensen, T., White, T., Wandall, H.H., Kristensen, A.K., Roepstorff, P., Clausen, H.: UDP-*N*-acetyl-α-D-galactosamine: polypeptide *N*-Acetylgalactosaminyltransferase. Identification and separation of two distinct transferase activities. J. Biol. Chem. **270**, 24186–24173 (1995)
24. Mulder, H., Spronk, B.A., Schachter, H., Neeleman, A.P., van den Eijnden, D.H., De Jong-Brink, M., Kamerling, J.P., Vliegenthart, J.F.: Identification of a novel UDP-GalNAc:GlcNAc β -R β1-4 N-acetylgalactosaminyltransferase from the albumen gland and connective tissue of the snail *Lymnaea stagnalis*. Eur. J. Biochem. **227**, 175–185 (1995)
25. Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., Xiong, Z., Que, H., Xie, Y., Holland, P.W., Paps, J., Zhu, Y., Wu, F., Chen, Y., Wang, J., Peng, C., Meng, J., Yang, L., Liu, J., Wen, B., Zhang, N., Huang, Z., Zhu, Q., Feng, Y., Mount, A., Hedgecock, D., Xu, Z., Liu, Y., Domazet-Lošo, T., Du, Y., Sun, X., Zhang, S., Liu, B., Cheng, P., Jiang, X., Li, J., Fan, D., Wang, W., Fu, W., Wang, T., Wang, B., Zhang, J., Peng, Z., Li, Y., Li, N., Wang, J., Chen, M., He, Y., Tan, F., Song, X., Zheng, Q., Huang, R., Yang, H., Du, X., Chen, L., Yang, M., Gaffney, P.M., Wang, S., Luo, L., She, Z., Ming, Y., Huang, W., Zhang, S., Huang, B., Zhang, Y., Qu, T., Ni, P., Miao, G., Wang, J., Wang, Q., Steinberg, C.E., Wang, H., Li, N., Qian, L., Zhang, G., Li, Y., Yang, H., Liu, X., Wang, J., Yin, Y., Wang, J.: The oyster genome reveals stress adaptation and complexity of shell formation. Nature **490**, 49–54 (2012)
26. Hagen, F.K., Hazes, B., Raffo, R., de Sa, D., Tabak, L.A.: Structure-function analysis of the UDP-N acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. Essential residues lie in a predicted active site cleft resembling a lactose repressor fold. J. Biol. Chem. **274**, 6797–6803 (1999)
27. Fritz, T.A., Hurley, J.H., Trinh, L.B., Shiloach, J., Tabak, L.A.: The beginnings of mucin biosynthesis: the crystal structure of UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-T1. Proc. Natl. Acad. Sci. U. S. A. **101**, 15307–15312 (2004)
28. Tenno, M., Toba, S., Kézdy, F.J., Elhammer, A.P., Kurosaka, A.: Identification of two cysteine residues involved in the binding of UDP-GalNAc to UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase 1 (GalNAc-T1). Eur. J. Biochem. **269**, 4308–4316 (2002)
29. Hassan, H., Reis, C.A., Bennett, E.P., Mirgorodskaya, E., Roepstorff, P., Hollingsworth, M.A., Burchell, J., Taylor-Papadimitriou, J., Clausen, H.: The lectin domain of UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities. J. Biol. Chem. **275**, 38197–38205 (2000)
30. Ten Hagen, K.G., Tran, D.T., Gerken, T.A., Stein, D.S., Zhang, Z.: Functional characterization and expression analysis of members of the UDP-GalNAc:polypeptide *N*-Acetylgalactosaminyltransferase family from *Drosophila melanogaster*. J. Biol. Chem. **278**, 35039–35048 (2003)