Universität für Bodenkultur Wien

Department of Chemistry

Division of Biochemistry

Head of Department: Ao. Univ. Prof. Dipl.-Ing. Dr. Erika Staudacher

Advisors: Ao. Univ. Prof. Dipl.-Ing. Dr. Erika Staudacher

Ao. Univ. Prof. Dipl.-Ing. Dr. Reingard Grabherr

ENZYMES INVOLVED IN GLYCAN PROCESSING IN GASTROPODS

Dissertation
for obtaining a doctorate degree
of the University of Natural Resources and Life Sciences
Vienna

Submitted by
Dipl. Ing. Chantal Lucini

Vienna, January 2014, Division of Biochemistry

Acknowledgements

The last three and a half years were an amazing job and life experience for me. Dr. Erika Staudacher gave me this great opportunity and I am infinitely grateful for that. She was also always there to encourage and support me when things did not work and always gave me helpful advices. Exactly the same is true for Dr. Reingard Grabherr. She was a great help especially in cell and molecular biology matters and never let me alone with my scientifical problems. Both made me love my job.

Very special thanks to Dr. Christopher Taus for hundreds of coffees, helpful scientific discussions and to be not only a good colleague but also a friend.

It is difficult to list all the other great people I worked with at the University of Natural Resources and Life Sciences, without forgetting someone but I will try. Thank you dear Glycobiology collegues: Dr. Fritz Altmann, Dr. Martin Dragosits, Dipl. Ing. Clemens Gruber, Dr. Alba Hykollari, Dipl. Ing. Simone Kurz, Mag. Laura Neumann, Dipl. Ing. Marita Preims, Dr. Katharina Paschinger, Dipl. Ing. Elisabeth Svehla, Mag. Andreas Thader, Dipl. Ing. Markus Windwarder, Mag. Katharina Wozny, Ing. Thomas Dalik, Daniel Maresch, Karin Polacsek and Dr. Dubravko Rendic. Thank you amazing people from Reingard's lab: Dr. Wolfgang Ernst, Stefan Gross, Dr. Stefan Heinl, Mag. Silvia Heiss, Dipl. Ing. Miriam Klausberger, Mag. Monika, Wilde, Dr. Dieter Palmerger, Dr. Jürgen Mairhofer, Mag. Diana Perez-Lopez, Dr. Katharina Spath, Christopher Tauer, Philipp Tauber, Dipl. Ing. Peter Vaboschek and Mag. Stella Wolfgruber. I had a great time in both labs and I am sorry if I forgot someone. I want also to thank Dr. Stefan Böhmdorfer, Mag. Donatella Tesei, Dr. David Reinhart, Dipl. Ing. Johanna Gottschamel and Dr. Sonja Gaier for the nice scientific and non-scientific chats.

I would also like to thank Dr. Regina Grillary for the helpful advices with the *Biomphalaria glabrata* cell line and Dr. Hans Marx and his team for helping us with the *Pichia pastoris* library.

Special thanks go also to my "Beratungsteam" Dr. Iain Wilson, Dr. Dietmar Haltrich and Dr. Richard Strasser for the precious hints and advices regarding our work.

Without the support of my family, all this would not have been possible. Thank you so much mum, dad and Alessandro! My "Viennese" family, Herbert and Elisabeth Schenter, also supported me a lot with good food and good mood.

Special thanks to Mag. Elena Gismano and Dipl. Ing. Alexandra Huber, two great scientists and friends, and also to all my other friends. Life is so funny with you!

Last but not least, thank you Stefan Schenter to go through all ups and downs with me, to support me wherever you are able to and to be such an amazing person on my side.

Zusammenfassung

Viele Gastropoden sind Zwischenwirte von Parasiten in Menschen und Tieren. Besonders Landschnecken sind in einigen Gebieten eine Plage und verursachen beträchtliche Schäden in der Landwirtschaft. Trotzdem sind wenige Daten über Enzyme, die in Schnecken Glykane modifizieren, vorhanden. Es wäre hilfreich einen tieferen Einblick in die Schnecken Glykobiologie zu gewinnen, da Strukturen und Mechanismen die nicht in höheren Tieren vorhanden sind als Werkzeug für die Schädlingsbekämpfung gezielt eingesetzt werden könnten.

Der Schwerpunkt diese Arbeit lag auf zwei Enzymen: einerseits, einer typischen tierischen N- Glykosyltransferase, der α-1,6-Fucosyltransferase, die eine Rolle bei vielen physiologischen und pathologischen Prozessen spielt; anderseits der Exoglykosidase ß-Galactosidase , die auch bei vielen biologischen Funktionen beteiligt ist und industriell eine Rolle spielt.

Am Beginn des Projekts waren nicht viele genetische Informationen über Schnecken vorhanden, da kein Genom oder Transkriptom bisher vollständig sequenziert wurde. Zur Identifizierung der α-1,6-Fucosyltransferase wurde doppelsträngige cDNA synthetisiert und für zwei Strategien verwendet: (1) Screening einer Expressions-cDNA-Bibliothek in Insektenzellen und (2) Homologiesuche mit degenerierten Primern. Die zweite Strategie war erfolgreich und etwa 90% der Enzymsequenz aus *Arion lusitanicus* wurde aufgedeckt. Die Untersuchung der Sequenz zeigte, dass die α -1 ,6- Fucosyltransferase aus *A. lusitanicus* typische Eigenschaften aus der Enzymfamilie enthält: es ist ein Typ zwei Golgi-Transmembranprotein mit einer GDP-Fucose-Bindungsstelle und der charakteristischen SRC Homology 3 (SH3)-Peptid-Bindungsdomäne. Eine lösliche Form des unvollständigen Proteins wurde erfolgreich in *Pichia pastoris* und in Insektenzellen exprimiert, aber es war nicht möglich, die Aktivität nachzuweisen. Zwei weitere α-1,6-Fucosyltransferasen aus Mollusken, von denen die vollständigen Sequenzen vorhanden waren, wurden auch exprimiert, aber die typischen Enzymaktivitätsnachweise haben nicht funktioniert. Möglicherweise haben diese Enzyme besondere Anforderungen oder es könnte ein Zeichen dafür sein, dass inaktive Isoformen identifiziert worden sind.

Eine ganz andere Strategie um die α-1,6-Fucosyltransferase zu isolieren ohne die Sequenz zu kennen ist, dieses aus Mikrosomen zu reinigen. Mikrosomen von der embryonalen Zelllinie aus *B. glabrata* und *A. lusitanicus* wurden präpariert und in den radioaktiven Enzymaktivitätsnachweise war teilweise Aktivität sichtbar. Dennoch waren die Ergebnisse nicht zufriedenstellend und die Mikrosomenpräparation erfordert noch Optimierungsarbeit.

Eine ß-Galactosidase aus *A. lusitanicus* wurde durch vier chromatographische Schritte gereinigt: hydrophobe Wechselwirkung-Chromatographie , Anionenaustauscher-Chromatographie, Größenausschluss -Chromatographie und Affinitätschromatographie mit Substratanaloga. Es ist gelungen, genügend Material zu sammeln, um eine Bande aus dem Coomassie Polyacrylamid Gel herauszuschneiden und durch Elektrospray Massenspektrometrie zu analysieren. Die Peptidsequenzen zeigen aber keinerlei Homologie zu Glycosidasen.

Stichwörter: Gastropoden, Mollusken, α-1,6-Fucosyltransferase, ß-Galactosidase

Abstract

Many gastropods are intermediate hosts of human and animal parasites. Especially land slugs are a pest in some areas and cause severe damages to agriculture. Nevertheless, scattered information is available on gastropod glycan modifying enzymes. To gain deeper insight into gastropod's glycobiology would be an important tool for pest control, as structures and mechanisms which are not present in higher animals could be targeted specifically.

In this work the focus was on two enzymes: firstly, a typical animal N-glycosyltransferase, α-1,6-fucosyltransferase, which plays a role in many physiological and pathological processes; secondly, the exoglycosidase ß-galactosidase, which is also involved in many biological functions and is an industrially relevant enzyme.

At the beginning of the project not so much genetic information from gastropods was available as no genome or transcriptome was fully sequenced so far. To identify the α-1,6-fucosyltransferase double stranded cDNA was synthesized and used for two strategies: (1) screening of an expression cDNA library in insect cells and (2) homology search with degenerate primers. The second strategy was successful and approximately 90% of the *Arion lusitanicus* enzyme sequence was revealed. Sequence analysis showed that *A. lusitanicus* α-1,6-fucosyltransferase contained sequence features typical for this enzyme family: it is a type two Golgi transmembrane protein with a GDP-fucose binding site and the characteristic SRC Homology 3 (SH3)- peptide binding domain. A soluble form of the incomplete protein was expressed successfully in *Pichia pastoris* and insect cells but it was not possible to prove its activity. Two mollusc α-1,6-fucosyltransferases, from which the complete sequences were available, have also been expressed but the typical enzyme activity assays did not work. This raises questions on eventual special requirements of the enzymes or it could be a sign that inactive isoforms have been identified.

A completely different strategy to isolate the α-1,6-fucosyltransferase without knowing its sequence was to purify it starting from microsomes. Microsome preparations of *B. glabrata* embryonic cell line and *A. lusitanicus* were performed and radioactive enzyme activity assays gave some hints that it might be some activity. Nevertheless, the results were not satisfying and much optimization work has to be done on the microsome preparation protocol.

ß-galactosidase from *A. lusitanicus* was purified using four chromatographic steps: hydrophobic interaction chromatography, anion exchange chromatography, size exclusion chromatography and substrate analogue affinity chromatography. It was possible to get enough material to excise a band from Coomassie polyacrylamide gel and analyze it through electrospray mass spectrometry. The peptide sequences showed no homology to any glycosidase.

Keywords: gastropods, molluscs, α-1,6-fucosyltransferase, ß-galactosidase

Table of contents

1 Introduction 10

1.1 Gastropods 10

1.1.1 *Biomphalaria glabrata* 10

1.1.2 *Arion lusitanicus* 10

1.2 Glycoconjugates 11

1.2.1 Biological role of protein glycosylation 11

1.2.2 N – Glycosylation 12

1.2.3 O – Glycosylation 14

1.2.4 Gastropod N- and O-glycosylation 15

1.3 Glycosyltransferases 17

1.3.1 Gastropod glycosyltransferases 18

1.4 Fucosylation and fucosyltransferases 19

1.4.1 α-1,6-fucosylation and –fucosyltransferases 20

1.5 ß-galactosidase 21

2 Aims 23

3 Materials and Methods 24

3.1 Synthesis and screening of cDNA expression libraries from gastropods 24

3.1.1 Cell culture and biological material 24

3.1.1.1 *Escherichia coli* culture 24

3.1.1.2 *Pichia pastoris* culture 24

3.1.1.3 Insect cell culture 24

3.1.1.4 *Biomphalaria glabrata* embryonic (Bge) cell culture 24

3.1.1.5 *Biomphalaria glabrata, Achatina fulica and Arion lusitanicus* 25

3.1.1.6 RNA purification 26

3.1.1.6.1 TRI Reagent purification 26

3.1.1.6.2 RNA Purification with RNAeasy Mini kit 26

3.1.2 Poly(A) purification 26

3.1.3 Vector preparation for expression libraries 27

3.1.3.1 Purification of pGAPZ B, pBacPAK8 and pENTR1a 27

3.1.3.2 Inverse PCR 27

3.1.4 cDNA Library construction 29

3.1.4.1 Double stranded cDNA synthesis and cloning 29

3.1.4.2 PCR screening 30
3.1.4.3 Storage 31
3.1.4.4 Sequencing 31
3.1.5 Construction of an expression library in *P. pastoris* cells 32
3.1.5.1 Vector preparation for *P. pastoris* cells transformation 32
3.1.5.2 Electrocompetent *P. pastoris* cells 32
3.1.5.3 Transformation 32
3.1.5.4 Storage 32
3.1.6 Construction of an expression library in *Sf*9 cells 33
3.1.6.1 Proof of concept with α-1,6-fucosyltransferase from *Mus musculus* 33
3.1.6.1.1 Cloning of mFUT8 and recombination in the engineered baculovirus genome 33
3.1.6.1.2 Creation of a baculovirus stock 33
3.1.6.1.3 FACS 33
3.1.6.1.4 Immunofluorescence microscopy 34
3.1.6.2 Cotransfection of the pBAKPak 8 library in *Sf*9 cells 34
3.1.6.3 Screening of the expression library in *Sf*9 cells 34
3.1.6.4 Cell sorting 34
3.1.6.5 FACS control of sorting after three rounds 35
3.1.6.6 PCR control of sorting after three rounds 35
3.2 Identification of FUT8 by PCR in directional cDNA libraries from gastropods 35
3.2.1 Alignment of different FUT8 sequences and selection of primers for homology search 35
3.2.2 cDNA Library construction with Infusion SMARTer cDNA Directional Library Construction Kit 37
3.2.3 PCR strategy to identify FUT8 37
3.2.3.1 Nested PCR 37
3.2.3.2 T/A- cloning and sequencing of the nested PCR fragments 38
3.2.4 Strategies to find the whole FUT8 sequence 39
3.2.4.1 3’ and 5’ Rapid amplification of cDNA ends (RACE) ready cDNA synthesis 39
3.2.4.2 5’RACE 39
3.2.4.3 Blunt end cloning and sequencing of the 5’RACE fragment 40
3.2.4.4 3’RACE 41

3.2.4.5 Colony blot 42
3.2.4.6 Southern blot 43
3.2.4.7 Inverse PCR on blunt end self ligated cDNA 43
3.2.4.8 Inverse PCR on sticky-end self ligated cDNA 44
3.2.4.9 FUT8-enrichment of the cDNA library for next-generation sequencing 45
3.3 Establishing an enzyme activity assay for FUT8 47
3.3.1 Preparation of the acceptor substrate for FUT8 47
3.3.1.1 Pronase digestion 47
3.3.1.2 Removal of the sialic acids 48
3.3.1.3 N-glycopeptidase A (PNGase A) digestion 48
3.3.1.4 Cation exchange chromatography 48
3.3.1.5 ß-galactosidase digestion 48
3.3.1.6 Quantification 48
3.3.1.7 Quality control by fluorescence labelling and determination on HPLC 48
3.3.2 Microsome preparation 49
3.3.3 Enzyme activity assays 49
3.4 Cloning and expression of a soluble incomplete form of *A. lusitanicus* FUT8 (amino acid residues 29 to 507) 50
3.4.1 Restriction of pVT-Bac 50
3.4.2 Restriction of pPICZα FlagHis1 50
3.4.3 Cloning FUT8 in pVT-Bac and pPICZα FlagHis1 51
3.4.4 Transformation of *P. pastoris* cells 52
3.4.5 *P. pastoris* cells PCR screening 53
3.4.6 Expression in *P. pastoris* cells 53
3.4.7 *Sf*9 Cotransfection with Baculogold DNA, infection and scale up. 53
3.4.8 SDS PAGE and Western Blot 54
3.5 Cloning and expression of four soluble FUT8 fusion proteins 55
3.5.1 Design of the constructs 55
3.5.2 Cloning of the constructs 55
3.6 Cloning and expression of *Lottia gigantea* and *Crassostrea gigas* FUT8s 57
3.6.1 Design 57
3.6.2 Cloning 57
3.7 Purification of ß- galactosidase from *A. lusitanicus* 58
3.7.1 Sample preparation and ammonium sulphate precipitation 58
3.7.2 Hydrophobic interaction chromatography 58

3.7.3 Anion exchange chromatography 58
3.7.4 Size exclusion chromatography 58
3.7.5 Affinity chromatography 58
3.7.6 SDS PAGE – silver staining 59
3.7.7 Enzyme activity assay 59
3.7.8 Mass spectrometry analysis of the purified ß-galactosidase 60
4 Results 61
4.1 Synthesis and screening of cDNA expression libraries from gastropods 61
4.1.1 Double stranded cDNA synthesis and cloning 62
4.1.2 Construction of an expression library in *P. pastoris* cells 64
4.1.3 Construction of an expression library in *Sf*9 cells 64
4.1.3.1 Proof of concept with α-1,6-fucosyltransferase from *Mus musculus* 64
4.1.3.1.1 Cloning of mFUT8, recombination in the engineered baculovirus genome, creation of the baculovirus stock 64
4.1.3.1.2 FACS 65
4.1.3.1.3 Immunofluorescence microscopy 66
4.1.3.2 Screening of the expression library in *Sf*9 by FACS 67
4.1.3.3 Cell sorting 67
4.1.3.4 FACS control of sorting after three rounds 69
4.1.3.5 PCR control of sorting after three rounds 69
4.2 Identification of FUT8 by PCR in directional cDNA libraries from gastropods 70
4.2.1 Alignment of different FUT8 sequences and selection of primers for homology search 70
4.2.2 cDNA Library construction 71
4.2.3 PCR strategy to identify FUT8 72
4.2.3.1 Nested PCR 72
4.2.3.2 T/A- cloning and sequencing of the nested PCR fragments 72
4.2.4 Strategies to find the whole FUT8 sequence 73
4.2.4.1 5'RACE 73
4.2.4.2 Blunt end cloning and sequencing of the 5'RACE fragment 74
4.2.4.3 3'RACE 75
4.2.4.4 Colony blot 76
4.2.4.5 Southern blot 76
4.2.4.6 Inverse PCR on blunt end self ligated and sticky end self ligated cDNA 77

4.2.4.7 FUT8-enrichment of the cDNA library for next-generation sequencing 77
4.3 Establishing an enzyme activity assay for FUT8 77
4.3.1 Preparation of the acceptor substrate for FUT8 78
4.3.2 Enzyme activity assay with homogenised rabbit brain 81
4.3.3 Microsome preparation of *A. lusitanicus* 82
4.3.4 Microsome preparation of Bge cells 83
4.4 Cloning and expression of a soluble incomplete form of *A. lusitanicus* FUT8 (amino acid residues 29 to 507) 83
4.4.1 Cloning FUT8 in pVT-Bac and pPICZα FlagHis1 84
4.4.2 Transformation and expression in *P. pastoris* cells 84
4.4.3 Cotransfection and expression of the incomplete FUT8 expressed in insect cells 86
4.5 Cloning and expression of four soluble FUT8 fusion proteins 88
4.5.1 Design of the constructs 88
4.5.2 Cloning of the constructs in pVT-Bac 89
4.6 Cloning and expression of *Lottia gigantea* and *Crassostrea gigas* FUT8s 91
4.7 ß- galactosidase purification from *A. lusitanicus* 92
4.7.1 Hydrophobic interaction chromatography 92
4.7.2 Anion exchange chromatography 93
4.7.3 Size exclusion chromatography 94
4.7.4 Affinity chromatography 95
4.7.5 SDS-PAGE with Coomassie and silver staining of the last chromatography step 97
5 Discussion 98
6 References 102
7 List of figures 111
8 List of tables 113
9 Table of abbreviations 114
10 Appendix 115
10.1 Glycan structures 115

1 Introduction

1.1 Gastropods

Gastropods, commonly known as snails and slugs, are the most highly diversified class in the phylum Mollusca, with approximately 60000 snail and slug species (Frýda et al, 2005): there are many thousands of species of sea snails and sea slugs, as well as freshwater snails, freshwater limpets, land snails and land slugs. They are only second to insect on their diversity at all taxonomic levels (McArthur & Harasewych, 2003). Snails are intermediate hosts of human and animal parasites, which cause infections of more than 200 million people worldwide and are a recognized veterinary problem in Africa and Asia (Hokke & Deelder, 2001; Vercruysse & Gabriel, 2005). The most common diseases transmitted by gastropods are schistosomiasis, clonorchiasis and paragonimiasis but these represent only a few of them with worldwide medical and economic impact. Especially land slugs are a pest in some areas and cause damage to vegetables, from small damages like bites on the leaves and tracks which reduce the commercial value, to complete crop failure.

To gain deeper insight in gastropod's glycobiology would be an important tool for pest control, as structures and mechanisms which are not present in higher animals could be targeted specifically. In our project we focused on the *Schistosoma* intermediate host *B. glabrata*, a freshwater snail, and on the vegetable devastating land slug *A. lusitanicus*.

1.1.1 *Biomphalaria glabrata*

B. glabrata is a species of air-breathing freshwater snails, an aquatic pulmonate gastropod mollusc in the family Planorbidae, the ram's horn snails. As this snail is considered a medically important pest, sequencing of the whole genome was approved as a priority by National Human Genome Research Institute in August 2004. Now, in 2013 on the homepage of the project http://biology.unm.edu/biomphalaria-genome/ it seems that annotation is underway. The genome length is 916,37 Mb which is a small genome size among gastropods (http://www.ncbi.nlm.nih.gov/genome/?term=txid6526[Organism:noexp]) and (http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/BiomphalariaSEQv.2.pdf).

In order to better understand the molecular genetic interactions between *Schistosoma mansoni* and the intermediate snail host *in vitro* the *B. glabrata embryonic* (Bge) cell line was established (Hansen, 1976). Until now it is the only established cell line from molluscs.

The *Schistosoma* life cycle is very complex: adult worm pairs live in the human mesenteric vessels, the female worm eggs are excreted with the feces and miracidia hatch from the eggs in freshwater reservoirs to infect the snail host. After asexual replication, the cercariae are released from sporocysts, which in turn penetrate the human skin, transform into schistosomula and develop into adult worms. It has been shown that the parasite expresses human-like sugar epitopes as molecular mimicry, such as Lewis X structures (Srivatsan et al, 1992) and it shares common carbohydrate epitopes with *B. glabrata*, namely a ß-1,2-linked xylose, a terminal α1,3FucGalNAc unit or both types of structural motifs (Lehr et al, 2007).

1.1.2 *Arion lusitanicus*

A. lusitanicus is a species of air-breathing land slugs, a terrestrial pulmonate gastropod mollusc of the family Arionidae, the roundback slugs.

It is considered among the 100 worst alien species in Europe in DAISIE European Invasive Alien Species Gateway (http://www.europe-aliens.org/) and it is the only land gastropod among these one hundred ones. *A. lusitanicus* is the worst slug pest in Europa and it has an important economical, ecological, health and social impact.

In order to find typical structures which could be a target for pest control, N- and O-glycans structures were elucidated (Gutternigg et al, 2004; Stepan et al, 2012). Those will be described in more detail in chapter 1.2.4.

1.2 Glycoconjugates

Glycoconjugates, or complex carbohydrates, are molecules in which one or more glycan units are covalently linked to a noncarbohydrate entity, either a lipid or a protein. There are different types of protein-linked glycosylation (

Figure 1):

- N-linked glycosylation: the sugar molecule is attached to a protein through the nitrogen atom of an asparagine residue.
- O-linked glycosylation: the sugar molecule is attached to a protein through the oxygen atom of an amino acid residue.
- Phospho-serine glycosylation: a sugar phosphate is transferred from a nucleotide sugar donor directly to a serine residue of a protein
- C-mannosylation: a mannose sugar is added to the first tryptophan residue in the sequence W-X-X-W
- Formation of glycophosphatidylinositol anchors (glypiation): a protein is attached to a lipid anchor via a glycan chain.

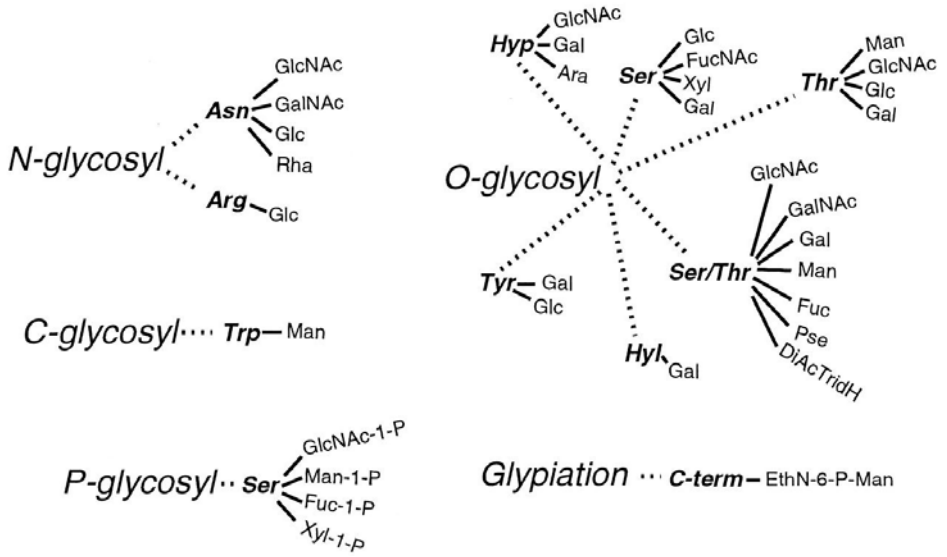


Figure 1: Types of protein-linked oligosaccharides (taken from Spiro, 2002).

There are also many different types of glycolipides, the most important ones are the glycosphingolipids and the glycoglycerolipids (Schnaar et al, 2009).

1.2.1 Biological role of protein glycosylation

Glycosylation is one of the most frequent and most important post-translational protein modifications. Their variety and complexity is unique: compaired to DNA, RNA, and proteins which are linear polymers, oligosaccharides can have branching and anomeric configurations (α and β linkages); whereas three amino acids or nucleotides can be combined into six possible sequences, three hexose monosaccharides can theoretically generate 1056 possible glycans. The biosynthesis of oligosaccharides is also extremely complex, not template-driven, varies among different cell types, and cannot be easily predicted from simple rules (Varki et al, 2009a).

The biological roles of glycans can be roughly divided into two categories (Varki, 1993; Varki & Lowe, 2009):

- Structural and modulatory:
  - Protection from proteases, shielding from antibodies, proper folding, solubility, conformation, charge…
  - Modulation of the interaction of proteins with one another, on-off/switching effects, tuning of primary functions of proteins…
- Specific recognition of glycans by other molecules (Elbein, 1987; Rademacher et al, 1988), most commonly, glycan-binding proteins, which can be subdivided into two major groups:
  - Glycan-binding proteins, which recognize glycans from the same organism: cell–cell recognition and cell–matrix interactions.
  - Glycan-binding proteins, which recognize glycans from a different organism: glycans as specific binding sites for pathogens, recognition targets for plant and bacterial toxins, molecular mimicry of pathogens, allergy epitopes (Tretter et al, 1993; Wilson et al, 1998).

As glycosylation has all these possible implications, care has to be taken when expressing a glycoprotein for therapeutic purposes in a non-human expression system. Different organisms have different enzymes and substrates which contribute to a different glycosylation pattern of the recombinant protein. In the worst case, this different or incorrect glycosylation can lead to complete loss of function or formation of allergenic structures (Aeed & Elhammer, 1994; Tretter et al, 1993). Therefore in the last years much effort has been put on glycoengineering of the N-glycosylation pathway of bacteria (Langdon et al, 2009), yeast (Wildt & Gerngross, 2005), plant (Loos & Steinkellner, 2012), insect (Tomiya et al, 2004) and mammalian cells (Durocher & Butler, 2009; Jacobs & Callewaert, 2009).

Glycans play also a role in different pathological conditions in humans, which can be roughly divided in the three following categories:

- Genetic glycosylation disorders: these rare human diseases are biochemically and clinically heterogeneous and usually affect multiple organ systems (Freeze & Schachter, 2009).
- Aquired diseases: several human disease conditions involve acquired (noninherited) changes in glycosylation and/or in the recognition of glycans (Freeze & Schachter, 2009).
- Cancer- altered glycosylation is a universal feature of cancer cells, and certain glycan structures are well-known markers for tumor progression (Varki et al, 2009b).

1.2.2 N – Glycosylation

N-glycans are covalently attached to a protein at the asparagine residue by an N-glycosidic bond. They have an N-acetylglucosamine (GlcNAc) linked to the asparagine.

In eukaryotic organism the synthesis starts on the cytoplasmatic side of the endoplasmatic reticulum (ER) membrane by transferring the GlcNAc-P from UDP-GlcNAc to the lipid-like precursor dolichol phosphate (Dol-P) to generate dolichol pyrophosphate N-acetylglucosamine (Dol-P-P-GlcNAc) and then the monosaccharides are added one by one by specif glycosyltransferases to the Dol-P until they reach the $Man_5GlcNAc_2$-PP-Dol structure. This structure is then “flipped” across the ER bilayer to the ER lumen. The mechanism is not fully understood and it is said to be mediated by a flippase, which seems to be related to the yeast Rtf1 locus (Helenius et al, 2002). Four mannoses and three glucoses are the added to the $Man_5GlcNAc_2$-PP-Dol structure until the N-glycan precursor $Glc_3Man_9GlcNAc_2$-P-P-Do is transferred *en bloc* by the oligosaccharyltransferase onto the growing polypeptide chain (Kornfeld & Kornfeld, 1985). The synthesis of the N-glycan precursor is shown in Figure 2.

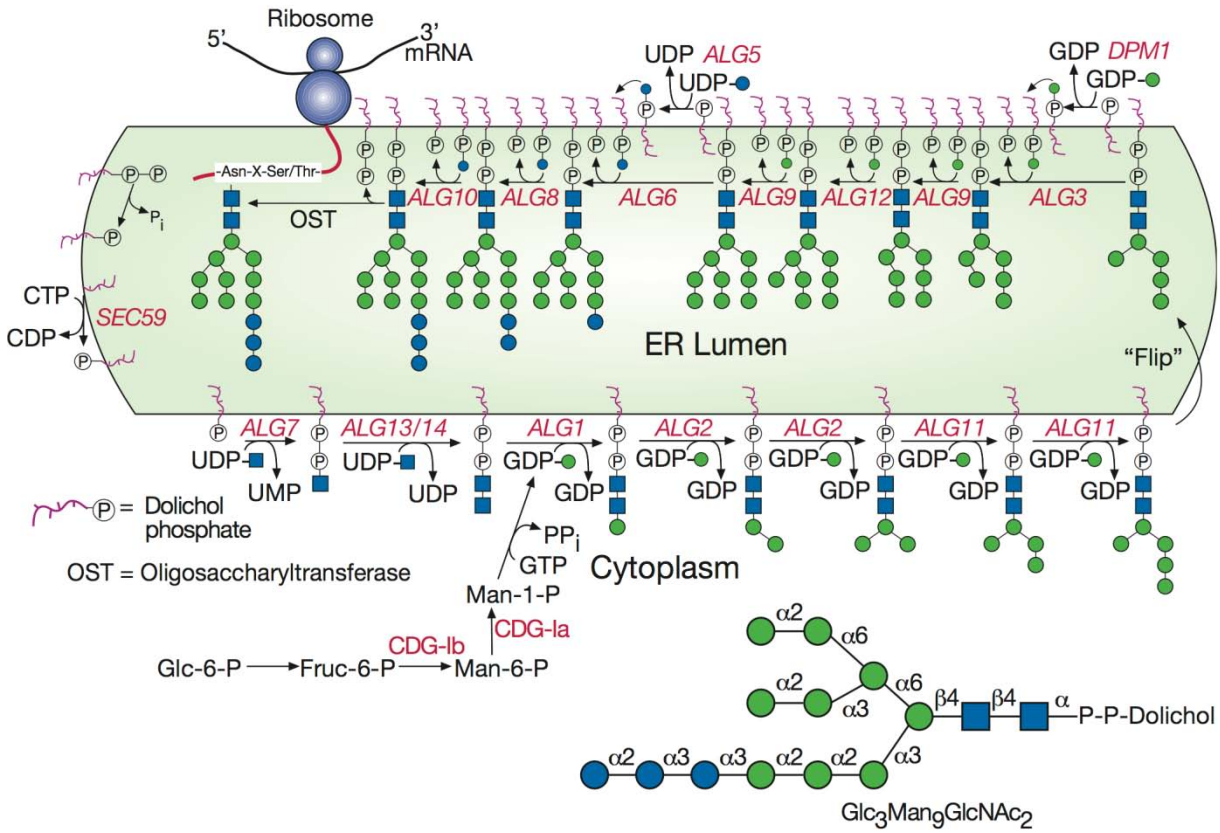


Figure 2: Synthesis of the N-glycan precursor (taken from Stanley et al, 2009).

The first steps of the processing of $Glc_3Man_9GlcNAc_2Asn$ are well conserved in eukaryotes and start with the sequential removal of glucose residues by α-glucosidases I (terminal glucose) and II (two inner glucoses). Often, an ER α-mannosidase I specifically removes the terminal α1–2Man from the central arm of $Man_9GlcNAc_2$ to yield a $Man_8GlcNAc_2$ isomer before the glycoprotein exits the ER. If properly folded, these proteins are packed in vesicles and transferred to the Golgi.

If the protein still contains the terminal glucose it binds to calnexin/calreticulin. This helps to ensure that proper protein folding occurred prior to exit from the ER. The proteins that are not folded properly are reglucosylated by the UDP-Glc:glycoprotein glucosyltransferase. When proteins completely fail to fold or oligomerize properly, they are often re-translocated into the cytoplasm and destroyed by N-deglycosylation and proteasomal degradation (Helenius, 1994; Helenius, 1997).

The further processing takes place in the Golgi: in multicellular eukaryotes three mannoses are often trimmed by the action of α-1,2 mannosidases IA, IB, and 1C in the *cis*-Golgi to give $Man_5GlcNAc_2$, a crucial intermediate for the synthesis of hybrid and complex N-glycans (Fuhrmann et al, 1985; Kornfeld & Kornfeld, 1985). Not all N-glycans are fully processed and some $Man_5GlcNAc_2$ escape further modification, therefore some secreted glycoprotein will carry oligomannose N-glycans of the type $Man_{5–9}GlcNAc_2$. Yeasts do not truncate the $Man_8GlcNAc_2$ N-glycans that enter the *cis*-Golgi and often add additional mannose residues to $Man_8GlcNAc_2$ to produce structures containing many branched mannose residues which are antigenic in human.

The synthesis of hybrid and complex glycans starts with the action of the N-acetylglucosaminyltransferase 1 (GlcNAcT-I, Figure 3), which transfers a GlcNAc residue to the Man-α-1,3 arm of $Man_5GlcNAc_2$. This reaction is essential for the action of all following enzymes which build hybrid or complex glycans: α-mannosidase II, GlcNAcT-II to V, α-1,6-fucosyltransferase (not in plants), β-1,2-xylosyltransferase (in plants, snails and trematodes) and core α-1,3-fucosyltransferase (in plants and invertebrates) (Schachter, 2000).

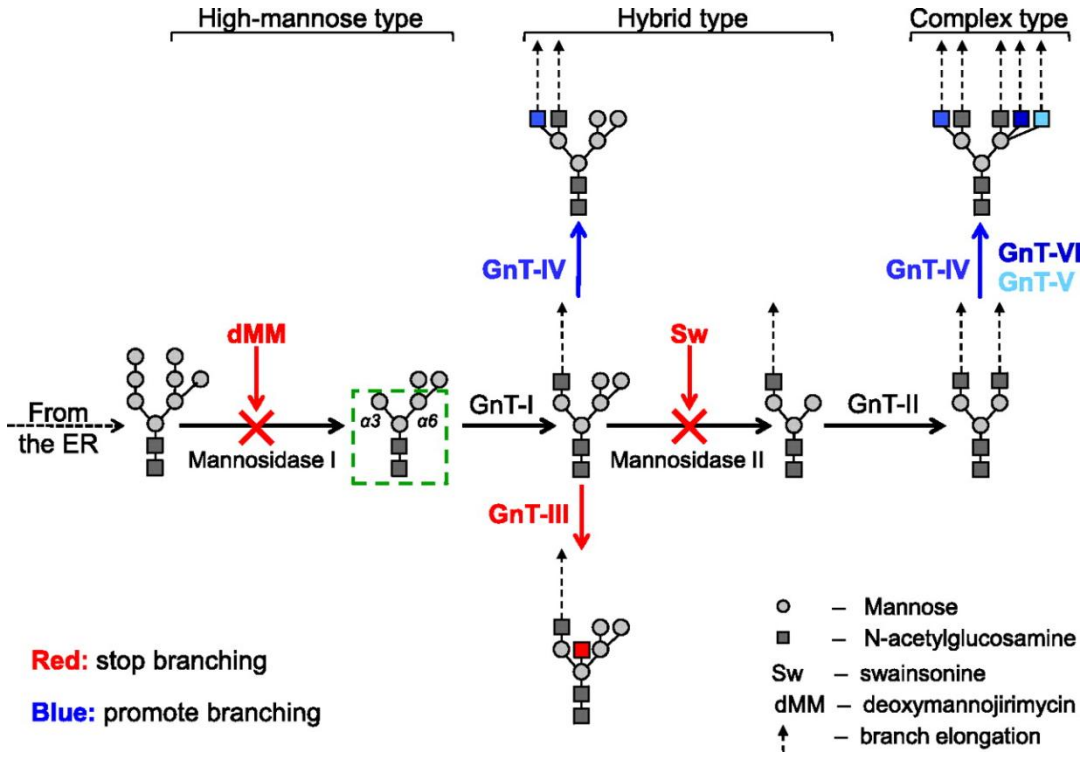


Figure 3: Synthesis of high mannose, hybrid and complex glycans (taken from Vagin et al, 2009).

The most common elongation of the branches of complex and hybrid N-glycans in higher eukaryotes starts with the addition of a β-linked galactose residue to the initiating N-acetylglucosamine and can be further lengthened by the sequential addition of N-acetylglucosamine and galactose residues. Sometimes, instead of β-linked galactose, β-linked *N*-acetylgalactosamine is added to N-acetylglucosamine. The branches are also often decorated with sialic acid, fucose, galactose, N-acetylgalactosamin and sulfate. A very rare modification is methylation, which has been found in the kingdom of animals only in worms and molluscs, whereas it is more frequently present in some species of bacteria, fungi, algae and plants, but not in mammals (Staudacher, 2012).

1.2.3 O – Glycosylation

O – glycans, in contrast to N – glycans, can be attached to different amino acids (serine, threonine, tyrosine, hydroxylysine, proline) with several sugar-residues such as GalNAc, N-acetylglucosamine (GlcNAc), mannose, glucose, galactose, arabinose, xylose, fucose, N – acetylfucosamine (Spiro, 2002). Depending on the linkage formed, the biosynthesis occurs in different organells.

The most frequent ones in eukaryotes are the mucin – type O – glycans. The first step of mucin O-glycosylation is the transfer of N-acetylgalactosamine from UDP-GalNAc to serine or threonine residues, which is catalyzed by a polypeptide-N-acetyl-galactosaminyltransferase (ppGalNAcT). In human at least 24 polypeptide ppGalNAcTs are found that differ in their amino acid sequences and are encoded by different genes (Ten

Hagen et al, 2003). Recently, the first gastropod ppGalNAcT from the water snail *B. glabrata* has been recombinantly expressed and characterized in our lab (Taus et al, 2013). With the addition of the next sugar, different mucin O-glycan core structures are synthesized (Figure 4)

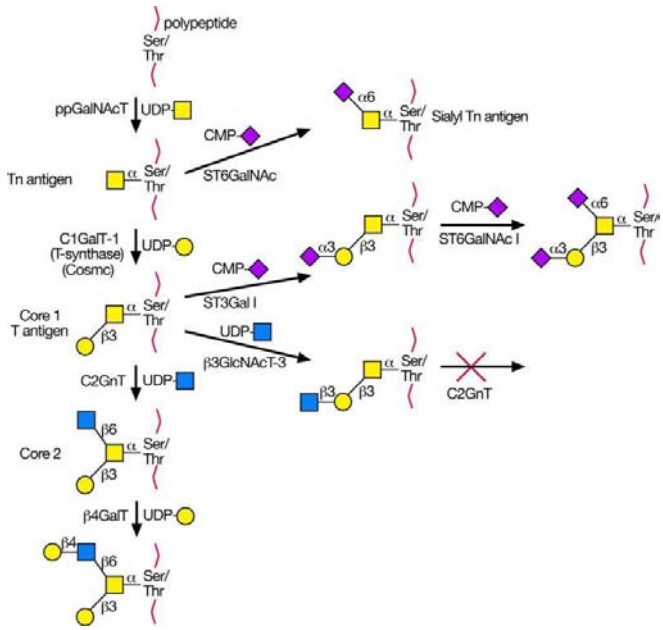


Figure 4: Biosynthesis of different mucin – O – glycan core structure (taken from Brockhausen et al, 2009).

Other structures commonly found in eukaryotes are GlcNAc- serine/threonine (Ser/Thr) without any elongation, a typical feature for nuclear and cytoskeletal proteins, and Fuc-Ser/Thr as well as Glc-Ser, which are found especially in epidermal growth factor (EGF) domains (Van den Steen et al, 1998). O-xylosylation is the starting point for the biosynthesis of chondroitin and heparan sulfates, which have roles in development and morphogenesis whereas O - mannosylation is the main O-glycan type in yeasts, but also some mammalian tissues carry this structural feature, which seems to be relevant in some muscular dystrophies (Lommel & Strahl, 2009). In plants O-glycans occur as arabinogalactans linked to hydroxyproline or serine or hydroxyproline is glycosylated with short arabinofuranosides (Spiro, 2002).

1.2.4 Gastropod N- and O-glycosylation

One of the first glycan characterizations in gastropods was on *Helix pomatia*'s hemocyanin, a copper – containing glycoprotein, which serves as an oxygen carrier in some Arthropoda and Mollusca (Ellerton et al, 1983). Dijk and Hall reported that it contained, in addition to fucose, xylose, mannose, galactose, glucose, GalNAc and GlcNAc an unidentified sugar residue (Dijk et al, 1970; Hall & Wood, 1976). One year later the unidentified sugar was confirmed to be a 3-O-methylgalactose (Hall et al, 1977). By elucidating the low-molecular weight structures, a ß-1,2-linked xylose to the ß-mannose was shown for the first time on an animal glycoprotein (Van Kuik et al, 1986; van Kuik et al, 1985) . Further, much more complex structures were found on *Helix pomatia*'s hemocyanin having a common core with an α-1,6-linked fucose on the reducing GlcNAc and a ß-1,2-linked xylose linked to the ß-mannose. One or both α-mannoses might be substituted with GalNAcß1,4GlcNAcß1,2-elements which contained two to four ß-1,3- or ß-1,6-linked galactoses with or without 3- or 4-O-methylgroups (Lommerse et al, 1997).

The low-molecular-mass N-linked carbohydrate chains of *Lymnea stagnalis* hemocyanin had common features with the ones in *Helix pomatia*: both species possessed a core structure with ß-1,2-linked xylose to the ß-mannose and they both have 3-O-methylated sugars, but

the 3-O-methylated mannose occurred only in *L. stagnalis* (Van Kuik et al, 1986) and no α-1,6-linked fucose on the reducing GlcNAc was found in *L. stagnalis* hemocyanin. High molecular mass structures also showed some antennae with the blood group H type 4 structure Fucαl,2Galß1,3GalNAcß (Van Kuik et al, 1987).

On the hemocyanin of the Californian giant keyhole limpet *Megathura crenulata* a number of the studied N-glycans contained a novel feature, namely a Galß1,6Man-unit (Kurokawa et al, 2002). In addition also N-glycans with an unusual elongation of two ß-1,4-linked galactoses to the core α-1,6 fucose were found (Wuhrer et al, 2004) as well as a terminal Fucα1,3GalNAc-motif, which is also an antigenic determinant in schistosomes (Geyer et al, 2004).

The primary structures of two biantennary N-glycans of the predatory sea snail *Rapana venosa*'s hemocyanin were also determined (Dolashka-Angelova et al, 2003) and the following structures proposed (Figure 5).

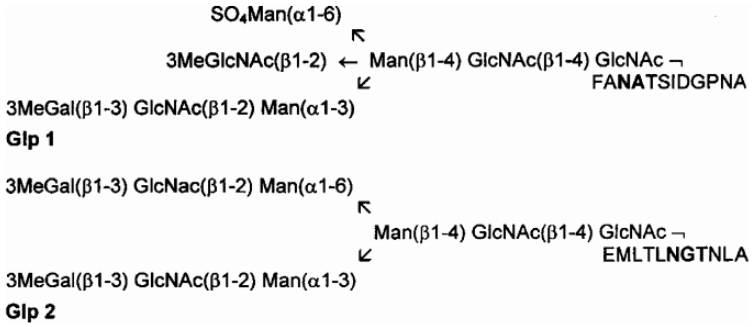


Figure 5: Two N-glycans of the functional unit of *Rapana venosa's* hemocyanin (taken from Dolashka-Angelova et al, 2003).

Further analysis on the same glycoprotein revealed also typical high-mannose glycans as well as an unusual fucose, which is attached to the GlcNAc residue located at the non-reducing end and is substituted at two positions with a hexosamine and a hexouronic acid (Sandra et al, 2007).

Gutternigg presented for the first time the neutral N-glycan structures of a whole gastropod, the slug *A. lusitanicus*, in two developmental stages (egg and adult)(Gutternigg et al, 2004). A wide range of structures were found:

- Oligomannosidic (five to nine mannoses) structures with or without 3-O-methylmannose.
- α-1,6-fucosylated structures which were always methylated in adult tissues.
- Methylated and non-methylated paucimannosidic structures (4 or less mannose residues) with additional core fucose (α-1,3- or α-1,6-linked) and/or xylose: here, in most cases the GlcNAc linked to the Manα1,3 was missing, probably due to a Golgi hexosaminidase trimming. Also, the core α-1,3-fucosyltransferase may not need or want any terminal GlcNAc as seen in *C. elegans* (Paschinger et al, 2004).
- Some complex structures with larger N-glycans with a number of galactose residues terminated with methyl groups, probably similar to those described by (Lommerse et al, 1997).

75% of the total N-glycans in eggs were unmethylated oligomannosidic structures, whereas in skin and viscera of adults they were just 8,8% and 47% respectively, but the most striking result was the complete absence of methylated structures in eggs.

In an older publication (Bürgmayr et al, 2001) the occurrence of N-acetylneuramnic and N-glycolyneuramnic in *A. lusitanicus* was proven. From these studies it can be said that *A. lusitanicus* has a big N-glycosylation potential as it combines structural features from mammals, plants, insects, nematodes and trematodes.

For the better understanding of the host-parasite interaction between *B. glabrata* and *Schistosoma mansoni*, the N-glycans of the snail cross-reacting with the trematode's

glycoconjugates were analyzed (Lehr et al, 2007). The cross-reacting species represented about 5% of the total glycans and exhibited a ß-1,2-linked xylose, a terminal Fucα1,3GalNAc unit or both types of structural motifs.

An extensive study on gastropod glycans can be found in (Gutternigg et al, 2007), where the neutral N-glycans of one water snail (*Planobarius corneus*), one slug (*Limax maximus*), two European (*Cepaea hortensis*, *Arianta arbustorum*) and one African (*Achatina fulica*) shell-carrying land species, were analyzed. Again, the most typical structural element found was the 3-O-methylation of terminal mannoses or galactoses. Most glycans were also α-1,6-fucosylated on the innermost GlcNAc, whereas α-1,3-fucosylation was quite rare. None of the glycans, as already seen in *A. lusitanicus* (Gutternigg et al, 2004), was difucosylated as it was found in insects (Staudacher et al, 1992). In *Planobarius* structures carrying a terminal α-1,2-fucose were found. A common modification in all analyzed species was the ß-1,2-linked xylose and also the lacking of the GlcNAc linked to the Manα1,3 was seen in almost 97% of the structures. Large structures were analyzed only in *Planobarius* and *Achatina*. The long 3-arm consisted of two N-acetylhexosamine residues linked together and was terminated by a methylated galactose. The short 6-arm carried only one or two mannoses substituted by a methyl group.

The first and only O-glycan analysis in gastropods was very recent (Stepan et al, 2012). Eight gastropod species, either land or water snails, with or without shell were analysed, namely *Achatina fulica*, *Arion lusitanicus*, *Biomphalaria glabrata*, *Cepaea hortensis*, *Clea helena*, *Helix pomatia*, *Limax maximus* and *Planobarius corneus*. The building blocks of the structures were just four monosaccharides: GalNAc, galactose, mannose and fucose. The only further modification is the methylation of mannose and galactose resulting in 3- or 4-O-Me-Gal and 3-O-Me-Man. Each O-glycan contained one amino sugar, which is the protein linked GalNAc. No other type of protein-linked sugar was detected. Recently the enzyme responsible for this linkage, the ppGalNAcT from the water snail *B. glabrata*, has been recombinantly expressed and characterized in our lab (Taus et al, 2013). The inner GalNAc is frequently elongated by two 4-O-Me-Gal residues in 1,3 and 1,6 linkages. Elongations of this core by one or two methylated or unmethylated hexoses appeared in most species, whereas further elongation was quite rare.

So far nothing is known about the biological functions of snail glycans and especially the methylation of hexoses in N- and O-glycans is quite unique and worth elucidation.

1.3 Glycosyltransferases

Glycosyltransferases (EC 2.4) are a large enzyme family which catalyzes the transfer of the monosaccharide moiety from a nucleotide sugar donor substrate onto an acceptor substrate, such as oligosaccharides, monosaccharides, polypeptides, lipids, small organic molecules, and even DNA. As they synthesize glycans, they play as well a role in many biological and pathological functions.

It is hypothesized that 1% of the open reading frames (ORF) of each genome is dedicated to the task of glycosidic bond synthesis (Coutinho et al, 2003). There are approximately 90 families of glycosyltransferases defined by primary structure analysis but they possess a rather limited number of fold types. All but a few can be either be classified as GT-A or GT-B folds glycosyltransferases (Figure 6). GT-A enzymes have a specific motif, the DXD or EXD motif, which is required for metal ion and donor substrate interaction. The folds of these enzymes contain a Rossmann fold of two tightly associated domains at the N-terminus. On the other hand, GT-B enzymes have folds consisting of two similar Rossmann folds (Coutinho et al, 2003; Qasba et al, 2005).

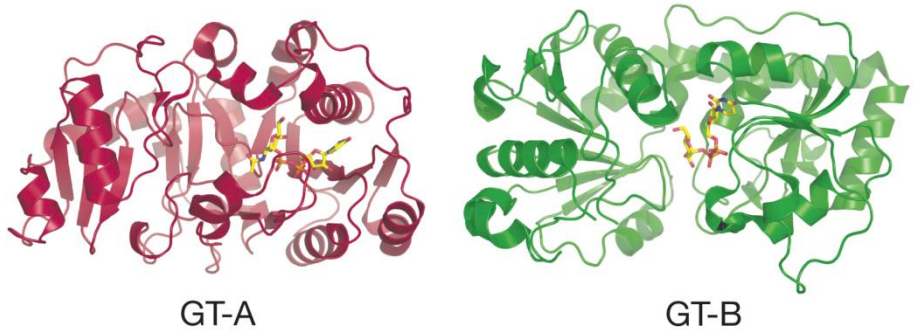


Figure 6: Ribbon diagrams of representative GT-A and GT-B folds (taken from Rini et al, 2009).

Glycosyltransferases are mostly localized in the Golgi apparatus. However, some of them are localized in ER, plasma membrane and/or outside of cells. Generally, glycosyltransferases involved in the synthesis of precursor structures or in early glycosylation steps tend to be in ER, and some are exceptionally localized in the plasma membrane (Furukawa et al, 2009). Most of the ER- and Golgi-resident glycosyltransferases are transmembrane proteins of various topologies, including single-pass-type membrane proteins with the catalytic globular domain outside the membrane, and integral membrane proteins with numerous transmembrane segments, the active site being formed by loops connecting some of the transmembrane segments. The type II topology is by far the most common protein architecture among the Golgi-resident GTs, consisting of a short N-terminal cytoplasmic tail followed by a trans-membrane domain, a stem region of variable length and a large C-terminal globular catalytic domain facing the luminal side (Hansen et al, 2010). A cleaving close to the membrane or an expression without transmembrane domain usually does not influence enzyme activity properties.

1.3.1 Gastropod glycosyltransferases

There are not many glycosyltransferases in gastropods that have been characterized so far and most of them are from the snail *Lymnea stagnalis*. Mulder showed that the connective tissue of *Lymnea stagnalis* contained ß-1,2-xylosyltransferase, GlcNAc-transferase I and GlcNAc-transferase II using a series of relevant substrates (Mulder et al, 1995a). Also a ß-1,4-GalNAc-transferase was identified by Mulder in the albumen gland and connective tissue of the same organism (Mulder et al, 1995b). Other enzymes identified in *Lymnea* were a ß-1,3-galactosyltatransferase (Mulder et al, 1991), an α-1,2-fucosyltransferase and a terminal α-1,3-fucosyltransferase (Mulder et al, 1991; Mulder et al, 1996). Also a core α-1,3-fucosyltransferase has been found (van Tetering et al, 1999), but no Lewis^{x}-structures nor core α-1,3-fucosylation have been detected in *Lymnea* so far. Using a cDNA probe of bovine ß-1,4-galactosyltransferase a novel glycosyltransferase was detected which belonged to the ß-1,4-galactosyltransferase gene family but used UDP-GlcNAc as sugar donor (Bakker et al, 1994). The acceptor substrate specificity was similar but much more restrictive than from the one of the bovine enzyme. In fact, this ß-1,4-GlcNAc-transferase is not involved in the synthesis of chitin-like molecules and both its polypeptide structure and acceptor specificity suggest that it neither is implicated in the synthesis of the chitobiose core of N-linked glycans. It is proposed that the enzyme functions in a novel, variant pathway of complex-type oligosaccharide synthesis in the snail (Bakker et al, 1997).

Another quite unique enzyme was identified in the albumen glands of the snail *Helix pomatia*: an α-1,2-L-galactosyltransferase which can use either GDP-L-galactose or GDP-L-fucose as donor sugar so that it is possible to synthesize blood group H active determinants (Lüttge et al, 1997). In *Biomphalaria glabrata*, *Helix pomatia* and *Arianta arbustorum* other galactosyltransferases were characterized which transfer D-galactose in linear chains or create branching points into galactans (Bretting et al, 2000; Stangier et al, 1995).

Recently a ppGalNAcT from *Biomphalaria glabrata* embryonic cells was identified by homology search in a cDNA library. The biochemical parameters of the enzyme expressed in insect cells confirmed a close relationship to the family of yet known ppGalNAcTs. The snail ppGalNAcT is highly homologous to the T2-enzymes from many other species and is capable to transfer GalNAc not only to specific blank polypeptides but also has follow-up activity on already glycosylated substrates (Taus et al, 2013).

1.4 Fucosylation and fucosyltransferases

Fucose is a monosaccharide that is found on glycoproteins and glycolipids in vertebrates, invertebrates, plants, and bacteria. Fucosyltransferases catalyze the inverting reaction in which a fucose residue is transferred from the donor GDP-fucose to the acceptor molecules including oligosaccharides, glycoproteins, and glycolipids. The fucosylated glycoconjugates are involved in a variety of biological and pathological processes.

Based on the site of fucose addition, fucosyltransferases are classified into α-1,2, α-1,3/4, α-1,6, and O-fucosyltransferase. α-1,2, α-1,3/4 and α-1,6 fucosyltransferases have probably evolved from one, or perhaps two, hypothetical ancestor gene(s), followed by duplications and subsequent divergence (Oriol et al, 1999).

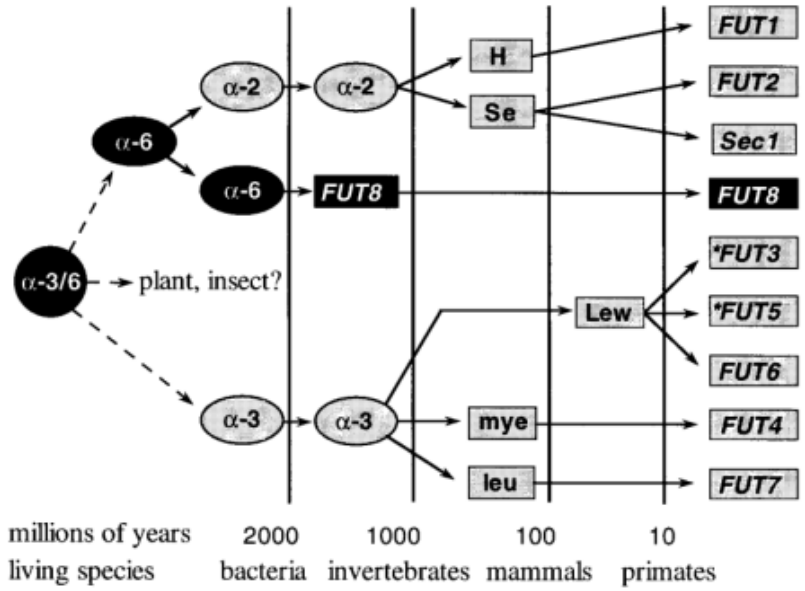


Figure 7: The hypothetical model of divergent evolution for the known fucosyltransferase genes (taken from Oriol et al, 1999). It is proposed that the fucosyltransferase gene family has a unique common ancestor (shown in the black circle). Probably, it used chitobiose as a substrate (black symbols) as it is expected from all members of the α-1,6-fucosyltransferase (FUT8) family. Plant and insect core α-1,3-fucosyltrasnferases may be the evolutionary link with other fucosyltransferases. The fucosyltransferases using N-acetyllactosamine as a substrate (grey symbols) have probably evolved later but share common peptid motifs (oval symbols). Now the fucosyltransferase genes have an overall sequence identity of over 30% (rectangular symbols).

These first three subfamilies of enzymes in eukaryotic organisms are type II transmembrane Golgi proteins containing an N-terminal cytoplasmic tail, a transmembrane domain, and an extended stem region followed by a large globular C-terminal catalytic domain facing the Golgi lumen. O- fucosyltransferases are endoplasmic reticulum-localized soluble proteins and catalyze O-fucosylation in the endoplasmic reticulum (Luo & Haltiwanger, 2005). An unusual non-Golgi α-1,2 fucosyltransferase has been found in the slime mold *Dictyostelium*: the enzyme is localized in both the cytoplasm and the nucleus. This enzyme lacks any conserved α-1,2 fucosyltransferase motifs (van Der Wel et al, 2001). Also *Helicobacter pylori* α-1,2 fucosyltransferase lacks the N-terminal cytosolic tail and the transmembrane domain and is a soluble protein located in the cytoplasm (Wang et al, 1999).

Fucose can be transferred directly to the hydroxyl group of serine and threonine residues of glycoprotein acceptors that contain either the EGF or the thrombospondin type repeat (TSR) and is there called O-fucose. The EGF-repeat sequence has been found in Notch, Notch ligands, urinary- and tissue-type plasminogen activators, and coagulation factors VII, IX, and VII, whereas the TSR sequence is present in extracellular matrix proteins involved in cell–cell and cell–matrix interactions (Ma et al, 2006).

In N-glycans and O-glycans, fucose is attached in α-1,2-linkage to galactose or to N- GlcNAc residues in 1,3, 1,4 or 1,6. α-1,3 and 1,6 linkage can be either in terminal position or on the N-glycan core, whereas the α-1,2- and 1,4-linkage have been found just in terminal position (Staudacher et al, 1999). Some few cases of elongations or branching of fucose have been described especially in molluscs (Gielens et al, 2005; Sandra et al, 2007; Wuhrer et al, 2004) and in *Schistosoma* (Khoo et al, 1997). In mammals fucosylated N-glycans have been found to be involved in sperm binding and in the course of development, changes in fucosylation activities can be observed in the various tissues of young animals. Fucosylation plays also a role in selectin mediated cell-adhesion and is often enhanced in cancer as well as in apoptosis (Staudacher et al, 1999). α-1,3-fucose bound to the proximal N-acetylglucosamine of an N-glycan as it is found in invertebrates and plants is a matter for both allergologists and biotechnologists because of its immunogenicity (Paschinger et al, 2005a; Tretter et al, 1993). Fucosylation plays also a role in many host-parasite interactions. For example in *Schistosoma mansoni* developmentally regulated expression of Le^x -structures were detected in O- and *N*-glycans of adult worms: these glycans induce changes in immune cell populations and the production of cytolytic autoantibodies in the vertebrate hosts (Staudacher et al, 1999).

1.4.1 α-1,6-fucosylation and –fucosyltransferases

Core α-1,6-fucosylation is a conserved feature of animal N-linked oligosaccharides being present in both invertebrates and vertebrates. The enzyme responsible for it, the α-1,6-fucosyltransferase (FUT8), catalyzes the transfer of fucose from GDP-fucose to N-linked type complex glycoproteins and belongs to the GT23 family of the CAZy classification. Both, the enzyme and the glycan modification, play an important role in several physiological and pathological processes. Nevertheless, α-1,6-fucosyltransferases from only nine organisms have been characterized (http://www.cazy.org/GT23_characterized.html), four of them belonging to *Rhyzobium* bacteria and five to eukaryotes, from which just two are from invertebrates. From the prokaryote *Bradyrhizobium* sp. WM9 (*Lupinus*) (Brzezinski et al, 2004) and human α-1,6-fucosyltransferase crystal structures exist. The structural analysis of a transmembrane domain-truncated form of the human α-1,6-fucosyltransferase showed that the enzyme consists of a catalytic domain, an N-terminal coiled-coil domain and a C-terminal SRC Homology 3 (SH3) domain. The catalytic domain was structurally classified as a member of the GT-B group of glycosyltransferases (Ihara et al, 2007).

It has been shown that *in vivo*, α-1,6-fucosylation protects glycans in humans against hydrolysis by glycoasparaginase (Noronkoski & Mononen, 1997) and is one necessary requirement for polysialylation (Kojima et al, 1996). Various studies using knockout mice also strongly suggested that α-1,6-fucosyltransferase and -fucosylation regulates especially receptor function, as the lack of core fucosylation perturbed the biological activities of various proteins (Takahashi et al, 2009; Taniguchi et al, 2006). In addition, core fucosylation was reported to be involved in antibody dependent cellular cytotoxicity. The lack of core fucose on N-glycans in the Fc region of the IgG1 molecule enhances antibody dependent cellular cytotoxicity activity up to 50-100-fold (Shields et al, 2002; Shinkawa et al, 2003). Therefore for the expression of antibodies for cancer therapy, *Fut8* knockout Chinese hamster ovary cells have been established (Yamane-Ohnuki et al, 2004). Highly α-1,6-fucosylated glycoproteins or FUT8 expression are also correlated with many diseases such as hepatocellular carcinoma (Hutchinson et al, 1991), ovarian serous carcinoma (Takahashi et al, 2000), papillary carcinoma of the thyroid (Ito et al, 2003), pancreatic cancer (Okuyama et

al, 2006), colorectal carcinoma (Muinelo-Romay et al, 2008) and cystic fibrosis (Wang et al, 1990).

The α-1,6-fucosyltransferase is quite strict on its acceptor requirements *in vivo* and *in vitro*. It has been shown that semi-purified mammalian, avian, and insect α-1,6-fucosyltransferases transfer fucose to biantennary oligosaccharides with nonreducing terminal N-acetylglucosamine residues (Longmore & Schachter, 1982; Staudacher et al, 1992; Struppe & Staudacher, 2000; Voynow et al, 1991; Wilson et al, 1976). They cannot use the core α-1,3-fucosylated substrate $GnGnF^3$ oligosaccharide (for the structure see chapter 10) (Staudacher & Marz, 1998). Also in *Caenorhabditis* α-1,6-fucosyltransferase requires prior action of N-acetylglucosaminyltransferase I. Unusually, the recombinant *Caenorhabditis* FUT8 can use a free $GnGnF^3$ oligosaccharide as a substrate. However, when using $GnGnF^3$ glycopeptide substrates (dabsyl- or dansyl-GnGn derived from fibrin or IgG and modified with *Arabidopsis* FucTA), both the native and recombinant *Caenorhabditis* FUT8 maintain the strict order (α-1,6 before α-1,3). This is probably due to the structure of the enzyme, which is shorter than the others characterized so far, and may therefore be more flexible when the free oligosaccharide is presented (Paschinger et al, 2005b). Especially for *in vitro* substrates it is necessary for the activity that the GlcNAc at the reducing end is intact, which reduces the possibilities of labelling.

1.5 ß-galactosidase

Glycosidases are extremely common enzymes with roles in degradation of biomass such as cellulose and hemicellulose, in anti-bacterial defense strategies (e.g., lysozyme), in pathogenesis mechanisms (e.g., viral neuraminidases) and in normal cellular function (e.g., trimming exoglycosidases involved in N-linked glycoprotein biosynthesis). Together with glycosyltransferases, glycosidases form the major catalytic machinery for the synthesis and breakage of glycosidic bonds. Glycosidases are divided in endoglycosidases, which cleave internal linkages in a glycosidic chain, releasing an oligosaccharidic residue and exoglycosidases, which cleave sugar chains at the outer (non-reducing) terminal end.

ß-galactosidase is an exoglycosidase which hydrolyzes the β-glycosidic bond formed between a galactose and the rest of the glycan. This enzyme is widespread in microorganisms, animals and plants. The one from *Escherichia coli* has been the most extensively studied and is valuable not only in the elucidation of the mechanism and specificity of glycosidase action but also for understanding genetic regulation of protein synthesis (Wallenfels & Weil, 1972). In mammalian organs and tissues, the widespread distribution of β-galactosidase is closely related to its numerous physiological functions. At least three different types of the enzyme can be identified according to their localization (Sinnott, 1990):

- The first type is located in the brush border membranes of enterocytes and is responsible for the hydrolysis of dietary lactose during digestion.
- The second type is located in lysosomes implicated in degradative processes. The existence of at least two genetically distinct acidic β-galactosidases has been reported in here (Kobayashi et al, 1985; Tanaka & Suzuki, 1977). Accumulation and overexpression of one of these endogeneous ß-galactosidases is a biomarker for senescent and aging cells (Dimri et al, 1995; Lee et al, 2006).
- The third type of β-galactosidase occurs in the cytosol.

Some non specific ß-galactosidases were purified from the digestive juice of the snails *Helix pomatia* (Got & Marnay, 1968) and *Achatina balteata* (Colas, 1980; Colas & Attias, 1977) and from the digestive gland of *Helicella ericetorum* (Calvo et al, 1983) and *Littorina littorea* (Cabezas et al, 1983): they possess not only β-galactosidase activity but also β-glucosidase and β-fucosidase activities that are associated with one or two active sites in one single protein.

A specific ß-galactosidase was purified from the digestive juice of *Achatina achatina*: it is soluble like the cytosolic β-galactosidases, functions at an acidic pH like the lysosomal

enzymes and it is strictly specific for the β-D-galactosyl residue. The purification was achieved by just three chromatographic steps: DEAE-Sepharose, Sephacryl S-200 and hydroxyapatite column. The purified enzyme is a large monomeric glycoprotein with a molecular mass (120–125 kDa). The amino acid composition displays a high amount of acidic/amide and hydroxy amino acid residues and a low content of basic residues. The enzyme activity is markedly affected by the ionic strength of the medium. This β-galactosidase is also capable of catalysing transgalactosylation reactions. The yields of galactosylation of hydroxy amino acid derivatives, catalysed by the enzyme in the presence of lactose as the glycosyl donor, were higher than those reported previously with conventional sources of β-galactosidases. In addition, the pH optimum is different for hydrolysis (pH 3.2) and transgalactosylation (pH 5.0) reactions (Leparoux et al, 1997).

2 Aims

The aim of this study was to identify, purify and characterize an α-1,6-fucosyltransferase from *Arion lusitanicus* and *Biomphalaria glabrata*. After the previous elucidation of N-glycan structures, we wanted to know the enzyme responsible for one of the most frequent modifications. In this work, three different strategies are described to isolate an α-1,6-fucosyltransferase without having any previous sequence information. The major challenges are the low level of expression of the enzyme, the presence of splicing variants and the enzyme activity.

We wanted also to purify a ß-galactosidase from *A. lusitanicus* to get enough material for electrospray mass spectrometry peptide analysis.

3 Materials and Methods

3.1 Synthesis and screening of cDNA expression libraries from gastropods

At the time this work was started no complete genome information was available for gastropods. It was necessary to create a cDNA library to clone in an expression vector for high-throughput screening with fluorescence-activated cell sorting (FACS).

3.1.1 Cell culture and biological material

3.1.1.1 *Escherichia coli* culture

If not otherwise described High Efficiency NEB 10-beta Competent *E. coli* cells (New England Biolabs, Frankfurt am Main, Germany) were always grown at 37°C overnight in Lysogeny Broth with the appropriate selection antibiotic (LB, 10 g/l peptone, 5 g/l yeast extract, 10 g/l NaCl in water) under constant shaking or on LB agar with the appropriate selection antibody in a 37°C incubator.

3.1.1.2 *Pichia pastoris* culture

If not otherwise described X-33 *P. pastoris* cells (Invitrogen, Vienna, Austria) were always grown at 30°C for 24 – 48 h min in Yeast Extract Peptone Dextrose (YPD, 10 g/l yeast extract, 20 g/l peptone, 20 g/l glucose in water**.** The glucose solution was autoclaved separately and added before use) containing 100 µg/ml Zeocin under constant shaking or on YPD agar (Add to YPD 20 g/l agar) containing 100 µg/ml Zeocin in a 30°C incubator.

3.1.1.3 Insect cell culture

If not otherwise described *Spodoptera frugiperda* cells (*Sf*9, ATCC CRL-1711), *Trichoplusia ni* High Five Cells (BTI-TN-5B1-4) or *Ascalapha odorata Ao*38 cells were cultivated in IPL41 medium (SAFC Biosciences, St. Louis, USA) containing yeast extract, a lipid mixture supplemented with 10% fetal calf serum (FCS), at 27°C using T-flasks (Summers et al, 1987).

3.1.1.4 *Biomphalaria glabrata* embryonic (Bge) cell culture

If not otherwise described embryonic cells from *Biomphalaria glabrata* (Bge cells, NR-40248, BEI Resources, NIAID, NIH) were grown as described at 26 °C using T-flasks (Knight et al, 2011). More in detail, Bge cells were ordered from BEI Resources, NIAID, NIH and arrived 3 days after they have been shipped from the USA. They arrived in culture in a roux which was completely filled with medium. The roux flask was incubated for a couple of hours at 26°C to allow the cells to attach, as they tend to detach during shipment. Then all but 10 ml medium were removed, the medium transferred to a sterile roux flask and stored at +4°C, in order that it could be used for adaptation of the cells to the new medium.

The medium was prepared as following:

Component	Company/Catalogue number	For 100 ml complete medium
Schneider's Drosophila Medium	Sigma-Aldrich #S9895	22 ml
Lactalbumin hydrolysate (sterile filtered stock 225 mg/ml)	Sigma-Aldrich #61302	2 ml
D-(+)-Galactose (sterile filtered stock 130 mg/ml)	Sigma-Aldrich #G5388	1 ml
Double distilled water		Add 20 ml first
Gentamycin (50 mg/ml stock)	Gibco, #15750-045	40 µl
Phenol red (0,5 % solution)	Sigma Aldrich #-P0290	162 µL

pH was adjusted to 7.0 with 1M NaOH, the volume was adjusted to 90 ml with double distilled water and the basic medium sterile filtered. For 100 ml complete medium 10 ml heat inactivated FCS (Gibco, FBS South American heat inactivated) were added. All pipetting was done under the hood, only the pH adjustment was done in the normal lab. It was important to not sterile filter FCS as some substances that were important for the cells might get lost.

Bge cells were splitted 1:2 or 1:3 once or twice a week depending on the actual grow rate. As the cells formed big dense floating cell aggregates, they were splitted by firm tapping and resuspended in fresh complete medium. They needed approximately 24 hours to reattach.

Confluent cells were frozen in freezing medium (90% heat inactivated FCS and 10% DMSO molecular biology grade). Therefore 900 µl freezing medium in 1,8 ml cryovials was placed on wet ice. The cells were detached by firm tapping and transferred to sterile falcon tubes. Then they were centrifuged for 5-10 min at 100-200 x g. The medium was removed and 900 µl of cold freezing medium was added dropwise. The cells were gently resuspended and transferred into cold cryovials with 900 µl freezing medium. The cells were kept in the wet ice box for 4 h at 4°C and then transferred in a styropor box at -80°C for at least 24 h or until completely frozen. Once completely frozen, they were quickly placed in liquid nitrogen.

To revive frozen cells from liquid nitrogen, the cells were quickly thawed to 80% in 35-37°C water bath (small piece of ice still remains) and then 1 ml of Bge basis medium (without FCS) at room temperature was added immediately. The cell suspension was transferred into a 15 ml tube, an additional 1 ml Bge basis medium was added and the cells were centrifuged at 700 rpm for 5 min at room temperature. The supernatant was discarded. The cell pellet was then resuspended in 11 ml Bge basis medium and centrifuged as above. The cells were finally resupended in 5 ml Bge complete medium (with FCS) and then transferred into a T25-tissue culture flask.

The cells were kept in culture at 26°C for a week without changing the medium to let cells adapt and start dividing. After a week, the whole medium was changed and the cells were grown for a further week. The medium was changed once a week until the cells were ready to split. When they were near confluence only half of the medium was changed. In sum, it took 4 weeks from thawing before the cells were ready to be splitted.

3.1.1.5 *Biomphalaria glabrata, Achatina fulica and Arion lusitanicus*

The snails *B. glabrata* and *A. fulica* were both bread in our lab whereas *A. lusitanicus* is a pest in our area and can easily be collected between May and September. For RNA preparation the living snails were put for 30 min at -80°C and then a piece of connective tissue cut with the scalpel. For other preparation the living snails were put at -20°C and stored there.

3.1.1.6 RNA purification

In principle, two methods were used to purify RNA: purification with TRI Reagent® (Sigma-Aldrich, Vienna, Austria) and RNeasy® Mini kit (Qiagen, Hilden, Germany).

3.1.1.6.1 TRI Reagent purification

For TRI Reagent purification approximately 50 mg fresh connective tissue from *B. glabrata*, *A. lusitanicus*, *A. fulica* or 10^7 *B. glabrata* embryonic (Bge) cells were homogenised in 1 ml TRI Reagent® with Ultra-turrax® (IKA, Staufen, Germany) at maximal speed three times for 10 sec. The samples were incubated for 5 min at room temperature and then centrifuged for 5 min at 11000 x g to remove cell debris. The supernatant was transferred into a fresh RNase free tube (Sarstedt, Wiener Neudorf, Austria) and 200 µl chloroform were added. The samples were vortexed vigorously for 15 sec, incubated for 3 min at room temperature and then centrifuged for 15 min at 4°C at 11000 x g. The colourless upper aqueous phase containing the RNA was transferred into a fresh RNAse free tube and then 500 µl isopropanol was added, the mixture was incubated for 10 min at room temperature and was centrifuged for 10 min at 4°C at 11000 x g. The supernatant was carefully removed and 1 ml 75% ethanol added to wash the RNA pellet. After vigorous vortexing, the sample was centrifuged at 7000 x g for 5 min at 4°C. The supernatant was then removed and the washing step repeated twice. The RNA pellet was finally resuspended in 100 µl diethylpyrocarbonate treated water (diethylpyrocarbonate was added to water to a final concentration of 0,1%, then the bottle was shaken vigorously and autoclaved) and the concentration was determined by measurement at 260 nm with Nanodrop 1000 UV-Vis spectrophotometer. The RNA was stored at -80°C.

3.1.1.6.2 RNA Purification with RNAeasy Mini kit

For purification with RNAeasy® Mini kit (Qiagen, Hilden, Germany) approximately 50 mg fresh tissue from *B. glabrata*, *A. lusitanicus*, *A. fulica* or $4x10^6$ Bge cells were homogenised in 350 µl buffer RLT (provided with the kit) containing 1% ß-mercaptoethanol using Ultra-turrax® at maximal speed three times for 10 seconds. Then RNA was purified according to the manufacturer's handbook provided with the kit. In short, 350 µl 70% ethanol were added to the homogenized sample and transferred to the RNeasy spin column provided with the kit. The sample was centrifuged 15 sec at 8000 x g and the flow-through discarded. 500 µl of buffer RPE (provided with the kit) were added to the column, centrifuged 15 sec at 8000 x g and the flow-through discarded. This washing step was repeated but with a longer centrifugation of 2 min at 8000 x g in order to dry the column. To elute the RNA, 30 µl RNAse free water (provided with the kit) were added to the column and centrifuged for 1 min at 8000 x g and collected in a clean RNase free tube. The RNA was stored at -80°C.

3.1.2 Poly(A) purification

In order to enrich the poly(A) RNA population, total RNA was further purified by either Dynabeads® Oligo $(dT)_{25}$ (Ambion, Vienna, Austria) or by MicroPoli(A) Purist (Ambion, Vienna, Austria). The starting material was approximately 75 µg of total RNA, which was processed according to the manufacturer's handbook. In short, for the Dynabeads® purification an equal volume of Binding Buffer (20 mM Tris/HCl pH 7.5, 1.0 M LiCl, 2 mM EDTA) was added to the total RNA solution. The mixture was heated at 65°C for 2 min and then was placed on ice. Then the RNA was added to the Dynabeads®, vortexed and rotated for 5 min at room temperature. The tube was then placed on the magnet and the supernatant was removed. 200 µl Washing Buffer B (10 mM Tris/HCl pH 7.5, 0.15 M LiCl, 1 mM EDTA) were added to the beads, mixed well by pipetting carefully and then the magnet was applied for 1 min. The supernatant was removed and the step was repeated once. To elute mRNA from the beads, Washing Buffer B was removed and 10 µl RNase-free water were added.

The RNA was incubated at 75°C for 2 min, and then the tube was placed on the magnet and the supernatant containing the mRNA was transferred to a new RNase-free tube. The mRNA was stored at -80°C.

3.1.3 Vector preparation for expression libraries

For the *P. pastoris* cells expression library the pGAPZ B vector (Invitrogen, Vienna, Austria) was taken, whereas the pBacPAK8 (Clontech, Saint-Germain-en-Laye, France) and pENTR1a (Invitrogen, Vienna, Austria) vectors were used for insect cell expression. The preparation consisted of two steps: first the plasmids were purified from an overnight culture. Then an inverse PCR was performed in order to linearize the plasmids and to add the complementary sequences which were necessary for recombination of the library with the plasmid.

3.1.3.1 Purification of pGAPZ B, pBacPAK8 and pENTR1a

For all plasmids, a 100 ml overnight culture was prepared. For pGAPZ B LB LS (LB low salt, the low salt variant of LB is used if the selection antibioticum is Zeocin. 10 g/l peptone, 5 g/l yeast extract, 5 g/l NaCl in water) contained 25 µg/ml Zeocin (Invitrogen, Vienna, Austria), whereas for pBacPAK8 LB with 100 µg/ml Ampicillin was used. For pENTR1a the final concentration of Kanamycin in LB was 50 µg/ml. All plasmids were purified by Nucleo Bond Xtra Midi plus kit (Machery-Nagel, Düren, Germany) according to the manufacturer's manual. In short: the cells were lysed and after neutralization the lysate was loaded on a column filter provided with the kit. Two washing steps were performed before elution, and then the plasmids were isopropanol precipitated by centrifugation at 16000 x g for 20 min at 4°C. The plasmid pellet was resuspended in 400 µl water and stored at -20°C.

3.1.3.2 Inverse PCR

The following primers were used for inverse PCR of the vectors:

Primer name	Sequence
pGAPZ B SMART forward	5' TTGATACCACTGCTTCCGCCAGCTTTCTAGAAC 3'
pGAPZ B SMART reverse	5' TTGATACCACTGCTTCCTCGTTTCGAAATAGTTG 3'
SMART_BAC8_f	5´TTGATACCACTGCTTGGCCTCGAGTTCGAATCTAG 3'
SMART_BAC8_r	5' TTGATACCACTGCTTTGCAGGGATCCGTATTTATAG 3
SMART_pENTR_noccdb_f	5´ TTGATACCACTGCTTATCTAGACCCAGCTTTCTT 3'
SMART_pENTR_noccdb_r	5' TTGATACCACTGCTTGGTTCCTTTAAAGCCTGCT 3'

For the inverse PCR of pGAPZ B Phusion High-Fidelity DNA Polymerase (New England Biolabs, Frankfurt am Main, Germany) was used.

The mastermix contained the following components:

Component	20 µl reaction
5x Phusion HF Buffer	4 µl
10 mM dNTPs	0,4 µl
10 µM Forward Primer	1 µl
10 µM Reverse Primer	1 µl
1 ng/µl pGAPZ B midi prep	0,8µl
Phusion DNA Polymerase	0,2 µl
H_2O	13,6 µl

The PCR started with an initial denaturation step at 98°C for 30 sec followed by 30 cycles with a denaturation step at 98°C for 10 sec, an annealing step at 57°C for 30 sec and an extension step at 72°C for 1 min 40 sec and a final extension step at 72°C for 5 min.

The linearized vector was run on an 1% agarose ethidium bromide gel electrophoresis and purified by gel extraction with NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany). First the band was excised from the gel and incubated at 50 °C with the buffer provided by the kit until completely dissolved. Then the solution was applied to the column provided with the kit and washed once. Elution was performed with 100 µl of water and then ethanol precipitation helped to get rid of the salts. Ethanol precipitation consisted in adding 1 µl of 10 mg/ml glycogen and 280 µl pure ethanol to the 100 µl solution. After a short incubation at -80°C and centrifugation at maximum speed for 30 min at room temperature, the pellet was finally re-dissolved in 10 µl water and stored at -20°C.

Also for the inverse PCR of pBacPAK8 Phusion High-Fidelity DNA Polymerase was used. The mastermix contained of the following components:

Component	20 µl reaction
5x Phusion HF Buffer	4 µl
10 mM dNTPs	0,8 µl
10 µM Forward Primer	1 µl
10 µM Reverse Primer	1 µl
1 ng/µl pBacPAK8 midi prep	4 µl
Phusion DNA Polymerase	0,2 µl
H_2O	10 µl

The PCR started with of an initial denaturation step at 98°C for 30 sec followed by 35 cycles with a denaturation step at 98°C for 10 sec , an annealing step at 60,9°C for 30 sec and an extension step at 72°C for 2 min and a final extension step at 72°C for 5 min.

The linearized pBacPAK8 was purified as previously described for pGAPZ B.

For pENTR1a nearly the same parameters apply:

Component	20 µl reaction
5x Phusion GC Buffer	4 µl
10 mM dNTPs	0,8 µl
10 µM Forward Primer	1 µl
10 µM Reverse Primer	1 µl
1 ng/µl pENTR1a midi prep	0,8 µl
Phusion DNA Polymerase	0,2 µl
H_2O	12,2 µl

The PCR started with of an initial denaturation step at 98°C for 30 sec followed by 35 cycles with a denaturation step at 98°C for 10 sec , an annealing step at 65°C for 30 sec and an extension step at 72°C for 1 min 20 sec and a final extension step at 72°C for 5 min.

3.1.4 cDNA Library construction

3.1.4.1 Double stranded cDNA synthesis and cloning

For the expression library it was necessary to synthesize double stranded cDNA (ds cDNA) first, which was afterwards cloned in the appropriate vector for further screening. In-Fusion SMARTer cDNA Library Construction Kit (Clontech, Saint-Germain-en-Laye, France) was used for all these steps. The starting material was approximately 100-200 ng of poly(A) RNA. First strand synthesis was performed by first mixing the RNA with the 3'SMART CDS Primer II A provided by the kit and adding H_2O to a final volume of 4,5 µl. The mix was incubated at 72°C for 3 min and then at 42°C for 2 min. In the second step of the first strand synthesis reverse transcription was performed by adding the following components:

Component	5,5 µl reaction
5x First strand buffer	2 µl
100 mM DTT	0,25 µl
10 mM dNTP	1 µl
12 µM SMARTer II A Oligonucleotide	1 µl
40 U/µl RNase Inhibitor	0,25 µl
100 U/µl SMARTScribe Reverse Transcriptase	1 µl

The mixture was incubated at 42°C for 90 min and the immediately placed on ice.

In order to synthesize the second strand the Advantage 2 PCR kit (Clontech, Saint-Germain-en-Laye, France) was used. The mastermix contained the following components:

Component	50 µl reaction
First strand cDNA	1 µl
H_2O	41 µl
10x Advantage 2 PCR Buffer	5 µl
10 mM dNTP	1 µl
12 µM 5'PCR Primer II A	1 µl
50x Advantage 2 Polymerase Mix	1 µl

The PCR started with an initial denaturation step at 95°C for 1 min followed by 27 cycles with a denaturation step at 95°C for 15 sec, an annealing step at 65°C for 30 sec and an extension step at 68°C for 6 min.

Ds cDNA was further purified and size-fractionated and then cloned into the linearized vectors by In-Fusion cloning. After optimization following conditions were used for the recombination reactions:

cDNA libraries	**pSMART2if**	**pGAPZ B**	**pBakPAK8**	**pENTR1a**
ds cDNA	750 ng	750 ng	200 ng	200 ng
Vector	300 ng	300 ng	100 ng	100 ng
5x Buffer	1x	1x	1x	1x
In-Fusion Enzyme	1 µl	1 µl	1 µl	1 µl
H_2O	to 10 µl	to 10 µl	to 10 µl	to 10 µl

3,3 µl of the recombination reactions were used for heat-shock transformation of 50 µl of High Efficiency NEB 10-beta Competent *E. coli* cells (New England Biolabs, Frankfurt am Main, Germany). The transformed *E. coli* cells were plated on LB agar with either 100 µg/ml Ampicillin (pSMARTif, pBacPAK8) or 50 µ/ml Kanamycin (pENTR1a) or LB LS agar with 25 µg/ml Zeocin (pGAPZ B) and incubated over night at 37°C.

3.1.4.2 PCR screening

In order to check the diversity of inserts cloned in the vectors PCR screenings of the colonies were performed. The following primers were used:

Primer name	**Sequence**	**Vector**
SMART forward	5’ CCTCTTCGCTATTACGCCAGC 3’	pSMART2if
SMART reverse	5’ TCACACAGGAAACAGCTATGA 3’	
- 44 back	5’ TTTACTGTTTTCGTAACAGTTTTG 3’	pBacPAK8
1660 for	5’ CAACGCACAGAATCTAGCGC 3’	
pGAP forward	5’ GTCCCTATTTCAATCAATTGAA 3’	pGAPZ B
3’ AOX	5’ GCAAATGGCATTCTGACATCC 3’	
pENTR1a_f	5’ GCAGGCTTTAAAGGAACC 3’	pENTR1a
pENTR1a_r	5’ CAAGAAAGCTGGGTCTAGAT 3’	

For the PCR screenings HybriPol DNA Polymerase (Bioline, London, UK) was used. Colonies were picked from the plates and resuspended in 50 µl H_2O. 5 µl were used as a template for PCR.

The mastermix contained the following components:

Component	30 µl reaction
10x Reaction Buffer	3 µl
10 mM dNTPs	0,6 µl
10 µM Forward Primer	0,3 µl
10 µM Reverse Primer	0,3 µl
50 mM $MgCl_2$	0,9 µl
Template	5 µl
HybriPol DNA Polymerase	0,3 µl
H_2O	19,6 µl

The PCR started with an initial denaturation step at 94°C for 5 min followed by 35 cycles with a denaturation step at 94°C for 15 sec, an annealing step at 55°C for 30 sec and an extension step at 72°C for 2min. The results were analyzed by gel electrophoresis on 1% agarose ethidium bromide gel.

3.1.4.3 Storage

One part of the colonies was washed out using 20 ml of LB medium (LB LS in case of pGAPZ B) for a big plate (25 x 25 cm). Aliquots of 1200 µl were taken and 100% glycerol was added to a final concentration of 25%. After vigorous mixing the libraries were stored at -80°C.

The other part of the colonies was also washed out and the plasmids purified by Nucleo Bond Xtra Midi plus kit (Machery-Nagel, Düren, Germany) to isolate the plasmids containing the library.

3.1.4.4 Sequencing

To check the diversity of the inserts and the presence of full ORFs from gastropod origin, overnight cultures of around 30 colonies were grown. The plasmids were purified by NucleoSpin Plasmid QuickPure (Machery-Nagel, Düren, Germany) and sequenced by MWG Eurofins (Vienna, Austria) with following primers:

Library	Primer
pSMART2if	SMART forward
pGAPZ B	pGAP forward
pBacPAK8	- 44 back

3.1.5 Construction of an expression library in *P. pastoris* cells

3.1.5.1 Vector preparation for *P. pastoris* cells transformation

As previously described in 3.1.4.3, plasmids were purified from 20 ml washed out colonies and used to transform electrocompetent X-33 *P. pastoris* cells.

3.1.5.2 Electrocompetent *P. pastoris* cells

Electrocompetent cells were made following this protocol:

- 5 ml overnight culture in YPD medium at 28°C was done.
- Next day the culture was used to set a 500 ml culture in YPD and let grow until OD 1,4 – 2,0 was reached.
- The cells were harvested at 2000 x g for 10 min at 4°C.
- The pellets were resuspended in 50 ml YPD prewarmed at 30°C. Then additional 50 ml YPD, 2 ml 1M HEPES pH 8.0 and 2,5 ml DTT were added and incubated at 30°C for 15 min.
- 400 ml ice-cold water were added and centrifugation was performed at 2000 x g for 10 min at 4°C.
- From this point it was worked on ice. The supernatant was discarded and the pellet resuspended in ice-cold 1 mM HEPES pH 6.0 (~250 ml).
- Centrifugation was performed at 2000 x g for 10 min at 4°C.
- The supernatant was discarded and the pellet resuspended in ~200 ml ice-cold 1M sorbitol.
- Centrifugation was performed at 2000 x g for 10 min at 4°C-
- The supernatant was discarded and the pellet resuspended in ~500 µl ice-cold 1M sorbitol
- 80 µl aliquots were frozen at -80°C.

3.1.5.3 Transformation

20 µl of the library plasmid preparation corresponding to 5 µg DNA (chapter 3.1.4.3) were used to transform 80 µl X-33 *P.pastoris* cells. The transformed yeasts were plated on YPD agar containing 100 µg/ml Zeocin and incubated at 30°C for 48 hours.

3.1.5.4 Storage

Storage of the yeast colonies containing the library was the same as previously described for *E. coli* cells except that the final glycerol concentration was 16%.

3.1.6 Construction of an expression library in *Sf*9 cells

3.1.6.1 Proof of concept with α-1,6-fucosyltransferase from *Mus musculus* (mFUT8)

In order to prove that it was possible to differentiate cells overexpressing α-1,6-fucosyltransferase from normal *Sf*9 cells which also express small amounts of FUT8, insect cells expressing the recombinant mFUT8 were compared to an infected control using FACS and immunofluorescence microscopy.

3.1.6.1.1 Cloning of mFUT8 and recombination in the engineered baculovirus genome

After purification of pENTR1a (chapter 3.1.3.1), the vector was digested with *XhoI* and *BamHI* HF (both from New England Biolabs, Frankfurt am Main, Germany). In parallel, starting from a plasmid containing mFUT8 (kind gift of Dr. Richard Strasser, Department of Applied Genetics und Cell Biology, BOKU), a two-step PCR with Phusion High-Fidelity DNA Polymerase (New England Biolabs, Frankfurt am Main, Germany) was performed using as a forward primer m_FUT8_XhoI_b (5' GATGATCTCGAGACTTCATCATACTTTTC 3') and mFUT8_f: 5' GATGATGGATCCCTCTGAAAATGCGGGCATG 3' as a reverse primer. The PCR product was digested with the respective enzymes and then ligated with the vector using T4 DNA ligase (New England Biolabs, Frankfurt am Main, Germany).

Transformation, PCR screening, mini prep and sequencing were performed as previously described in chapter 3.1.4.1 to 3.1.4.4. Plasmids of clones without mutations were purified, ethanol precipitated (chapter 3.1.3.2) and resuspended in 10 µl Tris EDTA (TE) buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA) pH 8.0. Recombination was performed using 2 µl LR Clonase (Invitrogen, Vienna, Austria), 150 ng linearized baculovirus genome attRcas, 150 ng plasmid preparation and TE buffer pH 8.0 to a final volume of 8 µl. The mix was incubated overnight at room temperature.

3.1.6.1.2 Creation of a baculovirus stock

8 µl of Cellfectin (Invitrogen, Vienna, Austria) were mixed into 100 µl IPL-41 (Tube A) and in a separate tube the 8 µl recombination reaction was added to 100 µl IPL41 (Tube B). Tube A was added to tube B and incubated for 30 min at room temperature. Then the mixture was put on $8*10^5$ *Sf*9 cells (without FCS) on a 6-well plate and incubated for 12 hours at room temperature. After removing the transfection medium, fresh medium containing 3% FCS and Penicillin/Streptomycin (PAA, Pasching, Austria) was given to the cells which incubated for 110 hours at 27°C.

The medium containing the virus was harvested and 80 µl of the suspension was used to infect $3*10^6$ *Sf*9 cells. After 110 hours at 27°C the medium containing the virus was again harvested and this suspension was then subjected to plaque assay in order to pick single plaques.

Single plaques were picked and propagated on 12-well plates, then PCR was performed under standard conditions (chapter 3.1.4.2) using pENTR_f and pENTR_r as primers) in order to confirm the presence of the mFUT8 PCR product. The supernatant of a positive plaque was used to further infect $3*10^6$ *Sf*9 cells for 110 hours at 27°C. The harvested virus suspension was used for FACS and immunofluorescence microscopy experiments.

3.1.6.1.3 FACS

For FACS analysis two wells of a 6-well plate *Sf*9 cells were infected: one with the baculovirus expressing mFUT8 and the other one with a baculovirus expressing a control protein.

Four days after infection cells were harvested and washed once with 1x phosphate buffer saline (PBS, PAA, Pasching, Austria). Then the cells were incubated under rotation at room

temperature for 1 hour with biotinylated *Lens culinaris* agglutinin (Vector Laboratories, Burlingame, USA) diluted 1:500 in 1x PBS with 0,5% FCS (blocking solution). After a washing step with PBS the cells were incubated with Streptavidin-Phytoerythrin (PE, Calbiochem (Merck), Vienna, Austria) 1:500 in blocking solution for 20 min under rotation at room temperature. The cells were then washed twice with PBS and resuspended in 400 µl fresh PBS before FACS analysis. FACS was performed using FACSCantoII at the Department of Nanotechnology.

3.1.6.1.4 Immunofluorescence microscopy

For immunofluorescence microscopy infection was done exactly as for the FACS analysis described above.

20 µl cell suspension were let adhere to a microscopy slide, washed with PBS and then fixed with 4% paraformaldehyd. PBS with 1% Triton was used to swell the cells and then a blocking step was performed using PBS with 20% FCS. The same lectin solution as for FACS analysis was added to the cells and incubated for 45 min. After washing with PBS, the Streptavidin-PE solution was applied for 20 min. One last washing step was performed and then the cells were covered with 50% glycerol. Immunofluorescence microscopy was performed using a Leica SP5 II laser scanning confocal microscope.

3.1.6.2 Cotransfection of the pBAKPak 8 library in *Sf*9 cells

45 µl of Cellfectin (Invitrogen, Vienna, Austria) were mixed into 455 µl IPL-41 (Tube A) and in a separate tube 20 µl cDNA library (corresponding to approx. 1 µg) were mixed with 12 µl 100 ng/µl Baculogold DNA (Clontech, Saint-Germain-en-Laye, France) and 81,5 µl IPL41 (Tube B). Tube A was added to tube B and incubated for 20 min at room temperature. Then the mixture was put on $9*10^5$ *Sf*9 cells (without FCS) on a 6-well plate and incubated for 4 h at room temperature. After removing the cotransfection medium, fresh medium containing 3% FCS and Penicillin/Streptomycin was given to the cells which incubated for 144 hours at 27°C. The medium containing the virus was harvested and used for all the following experiments.

3.1.6.3 Screening of the expression library in *Sf*9 cells

Before cell sorting, preliminary FACS experiments were done to eventually adjust the staining protocol. $9*10^5$ *Sf*9 cells were infected with 80 µl of the virus suspension containing the library. As negative control the same number of cells was infected with a baculovirus stock which expressed an arbitrary protein. The cells were incubated for 72 h at 27°C. The staining procedure was the same as described in chapter 3.1.6.1 but also *Aleuria aurantia* lectin (Vector Laboratories, Burlingame, USA) was tested.

3.1.6.4 Cell sorting

Cells overexpressing FUT8 should have a higher fluorescence signal when stained with *Lens culinaris* agglutinin. To enrich for those cells, FACS can be used to sort and collect the cells with the highest intensity.

For cell sorting the same protocol as described in chapter 3.1.6.1 was used, but $7,5*10^6$ cells were infected. Sorting was performed at with a FACSAria flow-cytometer (St. Anna Kinderspital). 5% of the cells with the highest fluorescence intensity were sorted and used to infect $7*10^5$ *Sf*9 cells for 7 d at 27°C. Then the supernatant was harvested and $7*10^5$ *Sf*9 cells infected with 80 µl of it. This virus was used in the following experiment. In parallel, the same number of cells was infected with the unsorted library to compare the intensity of the fluorescence signal by FACS. Finally $7,5*10^6$ cells were infected with the amplified sorted virus and the cells sorted. The whole procedure was repeated twice for a total of three rounds of sorting.

3.1.6.5 FACS control of sorting after three rounds

Five times 7,5*10^6 cells were infected with 80 µl virus stock: one with the original baculovirus library stock and the other three each with one round of sorting. The cells were stained and analysed as described in chapter 3.1.6.1.

3.1.6.6 PCR control of sorting after three rounds

PCR was performed under standard conditions (chapter 3.1.4.2) using pENTR_f and pENTR_r as primers.

3.2 Identification of FUT8 by PCR in directional cDNA libraries from gastropods

FUT8 sequences of different species show high homology at amino acid and nucleotide level. This feature was used as an additional strategy to fish the FUT8 out of the cDNA library.

3.2.1 Alignment of different FUT8 sequences and selection of primers for homology search

In order to design the appropriate degenerate primer the amino acid and the nucleotide sequences of FUT8s of different species were aligned. The sequences were obtained from http://www.ncbi.nlm.nih.gov/pubmed and different combinations of following species were used (gene accession number in parenthesis):

- **Primer set 1:** amino acid sequences of *Homo sapiens* (CAA76986.1), *Xenopus laevis* (AAI25985.1), *Drosophila melanogaster* (AAF48079.1) and *Caenorhabditis elegans* (CAD54736.1)

Primer name	Sequence
6FucT_f1	5' TGYCARYTICAYCAY 3'
6FucT_f2	5' GAYAARGTIGGIACIGAR 3'
6FucT_f3	5' YTIGTITGYACITTY 3'
6FucT_f4	5' GAYGAYATHTAYTAY 3'

- **Primer set 2:** nucleotide sequences of *Homo sapiens* (CAA76986.1), *Xenopus laevis* (AAI25985.1), *Drosophila melanogaster* (AAF48079.1) and *Caenorhabditis elegans* (CAD54736.1).

Primer name	Sequence
FUT8n1_f	5' GGSTGYCARCTBCAYCATGT 3'
FUT8n1_r	5'ACATGRTGVAGYTGRCASCC 3'
FUT8n2_f	5' AAAGTKGGMACVGAAGC 3'
FUT8n2_r	5' GCTTCBGTKCCMACTTT 3'

- **Primer set 3:** nucleotide sequences of *Lottia gigantea* from http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html (no GenBank submission), *Helobdiella robusta* from http://genome.jgi-psf.org/Helro1/Helro1.home.html (no GenBank submission) and *Ciona intestinalis* (CAD56161.1).

Primer name	Sequence
FUT8new_nt1_f	5' TTACARAAYCCHAAAGACTG 3'
FUT8new_nt1_r	5'CAGTCTTTDGGRTTYTGTAA 3'
FUT8new_nt2_f	5' TGTGGKTWYGGYTGTCA 3'
FUT8new_nt2_r	5' TGACARCCRWAMCCACA 3'
FUT8new_nt3_f	5' GTCTGGTGGRTKGGHCAG 3'
FUT8new_nt3_r	5'CTGDCCMAYCCACCAGAC 3'
FUT8new_nt4_f	5' TCYTCWCAGGTHTGTMG 3'
FUT8new_nt4_r	5' CKACADACCTGWGARGA 3'

- **Primer set 4:** amino acid sequences of *Homo sapiens* (CAA76986.1), *Xenopus laevis* (AAI25985.1), *Drosophila melanogaster* (AAF48079.1), *Caenorhabditis elegans* (CAD54736.1) and *Lottia gigantea* from http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html (no GenBank submission)

Primer name	Sequence
FUT8new_nt1_f	5' TTACARAAYCCHAAAGACTG 3'
AAFUT8_1_forward	5'GGITGYCARYTICAYCAYGTI 3'
AAFUT8_2_forward	5' MGIACIGAYAARGTIGGI 3'
AAFUT8_2_reverse	5' ICCIACYTTRTCIGTICK 3'
AAFUT8_3a_forward	5' YTIGTITGYACITTYWSIWSI 3'
AAFUT8_3b_forwardr	5'TTYWSIWSICARGTITGYMGI 3'
AAFUT8_3a_reverse	5' ISWISWRAAIGTRCAIACIAR 3'
AAFUT8_3b_reverser	5' ICKRCAIACYTGISWISWRAA 3'
AAFUT8_4_reverse	5' ICKRCAIACYTGISWISWRAA 3'
AAFUT8_2neu_forward	5' MGIACIGAYAARGTIGGIRYIGARGCIGCI 3'
AAFUT8_3neu_reverse	5' ICKRCAIACYTGISWISWRAAIGTRCAIACIAR 3'

- **Primer set 5:** nucleotide sequences of *Homo sapiens* (CAA76986.1), *Mus musculus* (AAH10666.1), *Drosophila melanogaster* (AAF48079.1) and *Caenorhabditis elegans* (CAD54736.1)

Primer name	Sequence
ntFUT8_1f	5' CAGGTVTGYCGVRTKG 3'
ntFUT8_2f	5'GAYGAYATMTACTAYT 3'
ntFUT8_2r	5' YCCAATGRTTWCCAGC 3'
ntFUT8_1r	5' CYTTRGAATWDCCATYC 3'

3.2.2 cDNA Library construction with Infusion SMARTer cDNA Directional Library Construction Kit

As the kit described in chapter 3.1.4 was not available anymore the newer version of it was used which had also the advantage that the ds cDNA was cloned in the correct direction.

The starting material was approximately 100-200 ng of poly(A) RNA. First strand synthesis was performed by first mixing the RNA with the 3' In-Fusion SMARTer CDS Primer provided by the kit and adding H_2O to a final volume of 4,5 µl. The mix was incubated at 72°C for 3 min and then at 42°C for 2 min. In the second step of the first strand synthesis reverse transcription was performed by adding the following components:

Component	**5,5 µl reaction**
5x First strand buffer	2 µl
100 mM DTT	0,25 µl
10 mM dNTP	1 µl
12 µM SMARTer V Oligonucleotide	1 µl
40 U/µl RNase Inhibitor	0,25 µl
100 U/µl SMARTScribe Reverse Transcriptase	1 µl

The mixture was incubated at 42°C for 90 min and at 68°C for 10 min.

In order to synthesize the second strand the Advantage® 2 PCR kit (Clontech, Saint-Germain-en-Laye, France) was used. The mastermix contained the following components:

Component	**50 µl reaction**
First strand cDNA	1 µl
H_2O	40 µl
10x Advantage 2 PCR Buffer	5 µl
10 mM dNTP	1 µl
12 µM 3' In-Fusion SMARTer PCR Primer	
12 µM 5'PCR Primer II A	1 µl
50x Advantage 2 Polymerase Mix	1 µl

The PCR started with an initial denaturation step at 95°C for 1 min followed by 17 cycles with a denaturation step at 95°C for 15 sec, an annealing step at 65°C for 30 sec and an extension step at 68°C for 6 min.

The ds cDNA was not further processed but stored at -20°C. For the homology PCR dilutions of 1 ng/µl were used.

3.2.3 PCR strategy to identify FUT8

3.2.3.1 Nested PCR

Nested PCR was apploed because it is a good method to amplify templates which are present in very low amount. It was divided in to step: the first amplification was done with one set of outer primers ("out") and generally no product was visible on 1% agarose. Using a very small amount of the first PCR as a template, the second amplification was done using an inner set of primers ("in").

All the nested homology PCRs were slight variations of the following protocol. The variable parameters were mainly the annealing temperature, the extension time and the cycle number.

First amplification:

Component	30 µl reaction
10x Reaction Buffer	3 µl
10 mM dNTPs	0,6 µl
10 µM Forward Primer “out”	7,6 µl
10 µM Reverse Primer “out”	7,6 µl
50 mM $MgCl_2$	0,9 µl
1 ng/µl Directional cDNA library	10 µl
HybriPol DNA Polymerase	0,3 µl

The PCR started with an initial denaturation step at 94°C for 3 min followed by 40 cycles with a denaturation step at 94°C for 30 sec, an annealing step at 45°C for 40 sec and an extension step at 72°C for 30 sec and a final extension step at 72°C for 5 min.

Second amplification:

Component	30 µl reaction
10x Reaction Buffer	3 µl
10 mM dNTPs	0,6 µl
10 µM Forward Primer “in”	7,6 µl
10 µM Reverse Primer “in”	7,6 µl
50 mM $MgCl_2$	0,9 µl
First amplification	1 µl
HybriPol DNA Polymerase	0,3 µl
H_2O	9 µl

The PCR started with an initial denaturation step at 94°C for 3 min followed by 40 cycles with a denaturation step at 94°C for 30 sec, an annealing step at 45°C for 40 sec and an extension step at 72°C for 30 sec and a final extension at 72°C for 5 min.

The results were analyzed by gel electrophoresis on 1% agarose ethidium bromide gel.

3.2.3.2 T/A- cloning and sequencing of the nested PCR fragments

DNA fragments with the appropriate size were cleaned up with NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany, see chapter 3.1.3.2) and cloned into pGEM-T vector (Promega, Fitchburg, USA) through T/A cloning. This cloning technique was possible as HybriPol DNA Polymerase produces A-overhangs. The ligation mixture was the following:

Component	10 µl reaction
50 ng/µl pGEM-T vector	1 µl
Homology search product	3 µl
2X Rapid Ligation Buffer (provided with the kit)	5 µl
50 U/µl T4 DNA Ligase	1 µl

The reaction was incubated for 2 h at room temperature.

3,3 µl of the ligation reaction was used for heat-shock transformation of 50 µl of High Efficiency NEB 10-beta Competent *E. coli* cells (New England Biolabs, Frankfurt am Main, Germany). The transformed *E. coli* cells were plated on LB agar with 100 µg/ml Ampicillin and incubated over night at 37°C. PCR screening (see chapter 3.1.4.2) was performed using T7 (5' TAATACGACTCACTATAGGG 3') and SP6 (5' ATTTAGGTGACACTATAG 3') primers. Overnight cultures of the positive clones were grown and the plasmids purified by NucleoSpin Plasmid QuickPure (Machery-Nagel, Düren, Germany). Sequencing was done by MWG Eurofins (Vienna, Austria) using the T7 primer.

3.2.4 Strategies to find the whole FUT8 sequence

3.2.4.1 3' and 5' Rapid amplification of cDNA ends (RACE) ready cDNA synthesis

5' and 3' Ready RACE cDNA were synthesised from 100 ng poly(A) RNA (see chapter 3.1.2) by using the SMARTer™ RACE cDNA Amplification Kit (Takara, Saint-Germain-en-Laye, France). The RNA was mixed with either the 3' or the 5' CDS Primer A (provided with the kit) and the volume adjusted with H_2O to 3,75 µl for 3' RACE and to 4,75 µl for 5' RACE. The mix was incubated at 72°C for 3 min and at 42°C for 2 min. To the 5' RACE reaction 1 µl of SMARTer IIA oligo was added. Then following mix was added to both RACE reactions:

Component	**5,25 µl reaction**
5x First strand buffer	2 µl
20 mM DTT	1µl
10 mM dNTP	1 µl
12 µM SMARTer V Oligonucleotide	1 µl
40 U/µl RNase Inhibitor	0,25 µl
100 U/µl SMARTScribe Reverse Transcriptase	1 µl

The tubes were incubated at 42°C for 90 min and 10 min at 70°C. For storage at -20°C and further experiments the RACE ready cDNA was diluted with 250 µL Tricine-EDTA Buffer (10 mM Tricine-KOH pH 8,5, 1 mM EDTA) provided with the kit.

3.2.4.2 5'RACE

To amplify the missing 5' end the two following primers were designed for nested PCR:

Primer name	**Sequence**
5RACE_FUT8Arion2a	5' CCAACTGTATCTCTAGGCAGAATATCC 3'
5RACE_FUT8Arion2b	5'CTGACAGGTCATGTGCATTCTGTCCTCC 3'

The forward primers are provided by the kit.

The mastermix of the first PCR contained of the following components:

Component	25 µl reaction
5x Phusion HF Buffer	5 µl
10 mM dNTPs	1 µl
10 µM 5RACE_FUT8Arion2a	1 µl
10x Universal Primer Mix	2,5 µl
5' RACE ready cDNA	2,5 µl
Phusion DNA Polymerase	0,2 µl
H_2O	12,8 µl

The PCR started with an initial denaturation step at 98°C for 30 sec followed by 30 cycles with a denaturation step at 98°C for 10 sec, an annealing step at 60°C for 30 sec and an extension step at 72°C for 1 min 20 sec and a final extension step at 72°C for 5 min.

The nested PCR contained of the following components:

Component	25 µl reaction
5x Phusion HF Buffer	5 µl
10 mM dNTPs	1 µl
10 µM 5RACE_FUT8Arion2b	1 µl
10 µM nested universal primer	1 µl
First PCR	1 µl
Phusion DNA Polymerase	0,2 µl
H_2O	15,8 µl

The PCR started with an initial denaturation step at 98°C for 30 sec followed by 30 cycles with a denaturation step at 98°C for 10 sec, an annealing step at 65°C for 30 sec and an extension step at 72°C for 1 min 20 sec and a final extension step at 72°C for 5 min.

3.2.4.3 Blunt end cloning and sequencing of the 5'RACE fragment

The 5'RACE DNA fragment was cleaned up with NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany, see chapter 3.1.3.2) and cloned by blunt end ligation into the pUC19 vector. Therefore, it was necessary to treat the DNA fragment with T4 polynucleotide kinase (T4 PNK, New England Biolabs, Frankfurt, Germany) as following:

Component	10 µL reaction
5' RACE DNA fragment (53,5 ng/µl)	5 µl
10x T4 Ligase Buffer	1 µl
10 U/µl T4 PNK	0,5 µl
H_2O	3,5 µl

The reaction was incubated for 30 min at 37°C and the PNK inactivated for 20 min at 65°C.

The ligation reaction was set as following:

Component	15 µL reaction
PNK reaction	10 µl
pUC19 (200 ng/µl)	0,5 µl
10x T4 Ligase Buffer	1 µl
400000 U/ml T4 DNA Ligase	1 µl
H_2O	2 µl

The reaction was incubated for 20 h at 16°C and 3,3 µl of it were used for heat-shock transformation of 50 µl of High Efficiency NEB 10-beta Competent *E. coli* cells (New England Biolabs, Frankfurt am Main, Germany). The transformed *E. coli* cells were plated on LB agar with 100 µg/ml Ampicillin and incubated over night at 37°C. PCR screening (see chapter 3.1.4.2) was performed using M13 forward (5' GTAAAACGACGGCCAGT 3') and M13 reverse (5' AACAGCTATGACCATG 3') primers. Overnight cultures of the positive clones were grown and the plasmids purified by NucleoSpin Plasmid QuickPure (Machery-Nagel, Düren, Germany). Sequencing was done by MWG Eurofins (Vienna, Austria) using the M13 forward primer.

3.2.4.4 3'RACE

Following primers were designed for 3'RACE:

Primer name	Sequence
3RACE_FUT8Arion	5' ATGGAGGACAGAATGCACATGACCTGTC 3'
FUT8Arion_f1	5' GGCTCATCATGGCAATACCGAGAACATG 3'
FUT8Arion_f2	5' GGATATTCTGCCTAGAGATACAGTTGG 3'
3RACE_FUT8Arion2a	5' GGAGGACAGAATGCACATGACCTGTCAG 3'
3RACE_FUT8Arion2b	5' GAGGCTCATCATGGCAATACCGAGAAC 3'
3RACE_FUT8Arion2c	5' ATCCTCGTCCGGGTTTTATGCCACT 3'
3RACE_FUT8Arion2d	5' CGTCGGTGTGGTGGATAGGTCACAT 3'
3RACE_FUT8Arion2e	5' CTTCCTGGTCTGCACGTTTTCATCG 3'
3RACE_Fut8_119	5' TTCCTCATGGCCTTCGGGACAATG 3'
3RACE_Fut8_414	5' GCTACGCAAGACAAACCATGCAGACG 3'
3RACE_Fut8_567	5' TGGCTGTCAGCTCCATCACATCACC 3'
3RACE_Fut8_837	5' GCCCGCAGATCTTGCTCATGACATC 3'
3RACE_Fut8_943	5' CAGGATGTGGTCAATGCTGGCAAGA 3'
3RACE_Fut8_1033	5' GCAGCCTTCCATCCCCTTTTGGAAT 3'
3RACE_1170	5' TGACCAGTTGGAGAGAACACAG 3'
3RACE_1202	5' CTAGGAGGGTATACCTGGCCTCAG 3'
3RACE_1298	5' CGCAGTCTGCATCGCTAGGTACTAG 3'
3RACE_1408	5' CAGGTTTGCAGAGTTGCCTATGAG 3'
3RACE_1470	5' CCGCTCACTTGATGACATCTTCTAC 3'
3RACE_1531	5' GAGGCCCATCATGGCAATACCGAG 3'
3RACE_1608	5' GGATATTCTGCCTGGAGATGCAGTTGG 3'

3FUT8_1211	5' TATACCTGGCCTCAGACGAC 3'
3FUT8_1321	5' TCTTGCTCGTTGCGACTTC 3'
3FUT8_1403	5' CATCGCAGGTTTGCAGAG 3'
3FUT8_1533	5' GGCCCATCATGGCAATACC 3'

All the following possibilities were tried for 3' RACE:

- Different primer combinations as outer and inner primer for nested PCR.
- Different annealing temperatures, extension times, and cycles number.
- Variations of the 5' RACE protocol described in 3.2.4.2 for Hybripol Polymerase (Bioline, London, UK) and OneTaq Polymerase (New England Biolabs, Frankfurt am Main, Germany).

Putative fragments were cloned either with blunt end cloning (chapter 3.2.4.4) or in case of fragments amplified with Hybripol and OneTaq with T/A cloning as described in chapter 3.2.3.2.

3.2.4.5 Colony blot

A colony blot of the non-directional cDNA library in pSMART2if (see chapter 3.1.4) was performed as following:

- **Plating of the colonies:** on a 576 cm^2 plate 10 ml of a 10-6 diluted cryostock (see chapter 3.1.4.3) were spread on a LB agar plate with 100 µg/ml Ampicillin and incubated overnight. The plated was then stored at 4°C
- **Synthesis of the probe:** using a plasmid preparation of chapter 3.2.4.3 as a template a standard PCR with Hybripol (see chapter 3.1.4.2) was performed using FUT8Hybrid_f (5' GACGACATATACTACTATGGAGGACAG 3') and FUT8Hybrid_r: (5' CCAACTGTATCTCTAGGCAGAATATCC 3') as primers. The PCR product was cleaned up with NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany, see chapter 3.1.3.2) and then labelled with digoxigenin (DIG) using the DIG HighPrime DNA Labelling and Detection Starter Kit I (Roche, Vienna, Austria). In short, approximately 1 µg of the amplified PCR fragment were adjusted to 16 µl with water, boiled for 10 min and quickly chilled on ice. Then 4 µl of the DIG-High Prime (provided with the kit) were added to the denatured DNA and incubated for 20 h at 37°C. The reaction was stopped by adding 0,2 M EDTA pH 8.0 and by heating at 65°C for 10 min. The probe was stored at -20°C.
- **Colony lift:** The positive charged nitrocellulose membrane (Roti-Nylon Plus, Roth, Karlsruhe, Germany) was carefully placed onto the pre-cooled agar plate. Then the membrane and plate were divided in four pieces and marked with a distinctive pattern. After 1 min the membrane was removed and the late incubated at RT for 1 d.
- **Washing and binding of the DNA to the membrane:** Firstly, the membranes were placed colony-side up onto filter paper saturated with 0.5 M NaOH/1.5 M NaCl and incubated for 5 min, then on a filter paper saturated with 1.5 M NaCl/0.5 M Tris/HCl, pH 7.4 and finally on a filter paper saturated with 2X saline-sodium citrate buffer (SSC, 20xSSC: this was the stock solution for the different diluition used in washing during southern and colony blotting. 3 M NaCl, 300 mM trisodium citrate pH 7.0). To bind the DNA to the membranes, they were baked for 30 min at 80°C.
- **Prehybridization and hybridization of the membrane:** DIG Easy Hyb (provided with the kit) was prewarmed to 37°C and was added to each of the four membrane pieces. The membranes were sealed in plastic bags and prehybridized for 30 min, shaking gently, at 37°C. In the meantime the DIG-labelled probe was denatured by boiling for 5 min and rapidly cooling in an ice/water bath. 5 µl of denatured DIG-labelled probe were

added to the preheated DIG Easy Hyb and added to the membranes. The membrane were sealed in a bag and incubated overnight at 37°C with gentle agitation.

- **Washes and detection:** The membranes were first washed for 5 min in 2X SSC, 0.1% SDS at room temperature with constant agitation and then twice 15 min in 0.5X SSC, 0.1% SDS (pre-warmed to wash temperature) at 62°C. The membranes were rinsed briefly in Washing Buffer (0,1 M maleic acid, 0,15 M NaCl pH 7.5, 0,3% Tween 20), incubated for 30 min in blocking solution first and then 30 min in Antibody solution (anti-digoxigenin-Alkaline Phosphatase provided with the kit diluted 1:5000 in blocking solution). Washing was performed twice for 15 min in washing buffer and then the membranes equilibrated for 5 min in Detection buffer (0,1 M Tris/HCl, 0,1M NaCl pH 9.5). Finally the membranes were incubated in freshly prepared Colour substrate solution (NCIB/BCIP stock solution provided with the kit diluted 1:50 in detection buffer) in the dark. The reaction was stopped by washing the membranes with water.
- **Sequencing:** Overnight cultures of the positive clones were set and the plasmids purified by mini prep with NucleoSpin Plasmid QuickPure (Machery-Nagel, Düren, Germany). Sequencing was done by MWG Eurofins (Vienna, Austria) using the SMART forward primer.

3.2.4.6 Southern blot

The directional cDNA synthesized as described in chapter 3.2.2 was cut near to the end with *MslI* (New England Biolabs, Frankfurt am Main, Germany) and the probe was designed to bind near to the 3' end. The experiment was done as following:

- **Restriction of the library and of the positive control:** 10 µg cDNA and 100 ng *Arion lusitanicus* FUT8 (sequence known to the date of the experiment) were digested with 10 U of MslI in a total reaction volume of 20 µl for 1 h at 37°C. The enzyme was then inactivated 30 min at 65°C.
- **Synthesis of the probe:** using a plasmid preparation of chapter 3.2.4.3 as a template a standard PCR with Hybripol (see chapter 3.1.4.2) was made using 3RACE_1408 and 5RACE_2a as primers. The DIG labelling was done exactly as described in chapter 3.2.4.5.
- **Membrane transfer:** A 1% agarose gel without ethidium bromide of the library and the positive control was run 1 h at 100 V. The transfer on a positive charged membrane (Roti-Nylon Plus, Roth, Karlsruhe, Germany) was performed as described in (Sambrook et al, 1989). In short: blotting buffer (0,4 M NaOH, 1 M NaCl) was poured into a 24x24 cm plate and then the blot built according to Sambrook (Sambrook et al, 1989). The transfer took place overnight at room temperature and on the next day DNA was fixed on the membrane by baking it at 120°C for 30 min.
- **Prehybridization and hybridization of the membrane:** this step was performed as described in chapter 3.2.4.5 but the prehybridization and hybridization temperature was set to 42°C.
- **Washes and detection:** this step was performed as described in chapter 3.2.4.5 but the membrane was stained with CPD Star (Roche, Vienna, Austria) working solution (CDP-Star provided with the kit diluted 1:100 in detection buffer) and visualized on the transilluminator Fusion FX7 (Vilber Lourmat, Eberhardzell, Germany) by exposing it 2, 15 and 30 min.

3.2.4.7 Inverse PCR on blunt end self ligated cDNA

3 µg of cDNA synthesized as described in chapter 3.2.2 were first ethanol precipitated (see chapter 3.1.3.2) and then treated with 5 U of DNA Polymerase I, Large (Klenow) Fragment (New England Biolabs, Frankfurt am Main, Germany) for 15 min at 25°C and then inactivated at 75°C for 20 min. The reaction was cleaned up with NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany, see chapter 3.1.3.2). 10 U of T4 PNK were added to it

and incubated for 30 min at 37°C. After 20 min inactivation at 65°C the cDNA was ligated as described in chapter 3.2.4.3.

For the inverse PCR the same primer listed in chapter 3.2.4.4 were used as forward primers, whereas following primers were used as reverse primer:

Primer name	Sequence
5FUT8_87	5' CGTTCTGCATGATTAGCTGCATCC '3
5FUT8_319	5' GCCTTTCTTCTGGCTTGTTCACCCTC '3
5FUT8_498	5' CTGCATGGTTTGTCTTGCGTAGCTTC '3
5FUT8_864	5' GGCTGTCAACGATAGGAAGGTCAATG '3
5FUT8_28	5' GCATAACCAGAAGGACAACAG '3
5FUT8_100	5' GACAAGCTACGTTCTGCATG '3
5FUT8_214	5' GTAACATTGTCCCGAAGGCC '3
5FUT8_214	5' CTGCATGGTTTGTCTTGCG '3

All the following possibilities were tried for inverse PCR:

- Different primer combinations as outer and inner primer for nested PCR.
- Different annealing temperatures, extension times, and cycles number.
- Variations of the 5' RACE protocol described in chapter 3.2.4.2 for Hybripol Polymerase (Bioline, London, UK) and OneTaq Polymerase (New England Biolabs, Frankfurt am Main, Germany).

Putative fragments were cloned either with blunt end cloning (chapter 3.2.4.3) or in case of fragments amplified with Hybripol and OneTaq with T/A clones as described in chapter 3.2.3.2.

3.2.4.8 Inverse PCR on sticky-end self ligated cDNA

To selfligate the cDNA it was necessary to synthesize a new ds cDNA which contains a restriction site which is also present in the FUT8 sequence. The enzyme of choice was *SacI* as it is a six base cutter which cuts the known sequence once approximately in the middle.

First strand synthesis was performed by first mixing 3 µg of RNA with SacI_3CDSPrimer (5' CGATGAGACACCAGAGCTCTTTTTTTTTTTTTTTTTTVN 3') and adding H_2O to a final volume of 4,5 µl. The mix was incubated at 72°C for 3 min and then at 42°C for 2 min. In the second step of the first strand synthesis reverse transcription was performed by adding the following components:

Component	5,5 µl reaction
5x First strand buffer	2 µl
100 mM DTT	0,25 µl
10 mM dNTP	1 µl
40 U/µl RNase Inhibitor	0,25 µl
100 U/µl SMARTScribe Reverse Transcriptase	1 µl
H_2O	1 µl

The mixture was incubated at 42°C for 90 min and at 68°C for 10 min.

In order to synthesize the second strand, the Advantage 2 PCR kit (Clontech, Saint-Germain-en-Laye, France) was used.

The mastermix contained the following components:

Component	50 µl reaction
First strand cDNA	1 µl
H_2O	40 µl
10x Advantage 2 PCR Buffer	5 µl
10 mM dNTP	1 µl
12 µM SacI_3CDSPrimer	1 µl
12 µM SacI_FUT8 5' GTGGAAGAAATGAGCTCGTATG 3'	1 µl
50x Advantage 2 Polymerase Mix	1 µl

The PCR started with an initial denaturation step at 95°C for 1 min followed by 17 cycles with a denaturation step at 95°C for 15 sec, an annealing step at 65°C for 30 sec and an extension step at 68°C for 6 min.

The ds cDNA was then cut with 10 U of *SacI* (New England Biolabs, Frankfurt am Main, Germany) for 2 h at 37°C and cleaned up with NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany, see chapter 3.1.3.2).The DNA was then finally ligated with T4 ligase as described in chapter 3.2.4.3.

Inverse PCR was performed as described in the previous chapter.

Putative fragments were cloned either by blunt end cloning (chapter 3.2.4.3) or in case of fragments amplified with Hybripol and OneTaq by T/A cloninig as described in chapter 3.2.3.2.

3.2.4.9 FUT8-enrichment of the cDNA library for next-generation sequencing

One possible approach to get the whole sequence was to enrich the cDNA for FUT8 and sequence it with a next-generation sequencing approach. The first step was to design one set of forward primers through the whole sequence coupled with biotin and a 15 carbon atom spacer (biotin-TEG) on the 3' end and a second pair of unlabelled reverse primers in order to build approximately 200 bp long probes. The biotinylated primers were also used alone.

Following primers were designed:

Code	Primer	Sequence
1	FutBiot1-30	3' biotin-TEG-ATGAAACAATGGAAAGTAATAGTATTACTG 5'
2	FutBiot400-429	3' biotin-TEG-GACCACATAGACAACAATGATAATCATTTG 5'
3	FutBiot710-738	3' biotin-TEG-GAACTCTGATACTTGATTCCAAAGGCTGG 5'
4	FutBiot1061-1090	3' biotin-TEG-GATTCCAAAACACCATTGTTGGGGTACATG 5'
5	FutBiot1432-1460-	3' biotin-TEG-ATTATGCAGACATTGCATGGAGATGCATC 5'
A	3>5Fut285-265	3' CTGAGACAGTTCATTGCTAGC 5'
B	3>5Fut647-625	3' CATCCTTTGTGCAAGTTGCAAAC 5'
C	3>5Fut948-928	3' GATGTCATGAGCAAGATCTG 5'
D	3>5Fut1267-1246	3' CTGGATATTGTTTCTGGGCCTC 5'
E	3>5Fut1608-1588	3' CCAATGGTTTCCAGCTATACC 5'

In order to synthesize the probes, a PCR with following primer combinations was done:

Forward primer (code)	Reverse primer (code)
1	A
2	B
3	C
4	D
5	E

The PCR parameters were the following:

Components	25 µl reaction
5x OneTaq® Standard Reaction Buffer	5 µl
10 mM dNTP	0,5 µl
Primer forward	0,5 µl
Primer reverse	0,5 µl
Mini prep pPIC-FUT8 1 ng/µl	1
Onetaq Polymerase	0,125
H_2O	22,375

The PCR started with an initial denaturation step at 94°C for 30 sec followed by 35 cycles with a denaturation step at 94°C for 30 sec, an annealing step at 60°C for 40 sec and an extension step at 68°C for 30 sec and a final extension step of 5 min. The PCR products were purified by gel extraction with NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany, see chapter 3.1.3.2)

The protocol for hybridization was a slightly modified version of (Li et al, 2011): the biotinylated probes (primers or PCR products) were added to 10 µl cDNA and the sample diluted with 6x SSPE (900 mM NaCl, 60 mM $NaH_2PO_4xH_2O$, 60 mM Na_2EDTA) to have a final concentration of each probe of 33 nmol/l in a total volume of 45 µl. The mixture was denatured at 100°C for 2 min in a PCR thermocycler and then quickly cooled on ice for 5 min. The hybridization was performed overnight at 58°C in a thermocycler. In order to get rid of the probe excess, the mixture was purified by NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany) as described in chapter 3.1.3.2 and eluted with 40 µl H_2O.

In the meantime 10 µl Dynabeads® M-280 Streptavidin (Invitrogen, Vienna, Austria) were washed twice with 1x B&W buffer (stock solution: 2x B&W buffer: 10 mM Tris/HCl pH 7.5, 1 mM EDTA, 2 M NaCl) and resuspended in 40 µl 2x B&W buffer + 0,1% Triton. 40 µl hybridization mixture was added to it and incubated under continuous rotation at room temperature for 1 h. After washing it three times with 1x B&W buffer the beads coupled with the DNA were resuspended in 15 µl water. In order to elute the DNA the beads were boiled for 5 min at 99°C and the solution immediately transferred in a clean tube.

As a full double strand is needed for further sequencing, the DNA was amplified using the same protocol as for double strand synthesis in the directional library construction:

Component	50 µl reaction
DNA eluted from beads	15 µl
H_2O	26 µl
10x Advantage 2 PCR Buffer	5 µl
10 mM dNTP	1 µl
12 µM 3' In-Fusion SMARTer PCR Primer	1 µl
12 µM 5'PCR Primer II A	1 µl
50x Advantage 2 Polymerase Mix	1 µl

The PCR started with an initial denaturation step at 95°C for 1 min followed by 15 cycles with a denaturation step at 95°C for 15 sec, an annealing step at 65°C for 30 sec and an extension step at 68°C for 2 min.

To check if enrichment was successful a control PCR was performed using two FUT8 sequence-specific primers and 35 cycles of amplification:

Components	25 µl reaction
5x OneTaq® Standard Reaction Buffer	5 µl
10 mM dNTP	0,5 µl
3RACE_119	0,5 µl
5RACE_Arion	0,5 µl
ds DNA from previous PCR	1 µl
Onetaq Polymerase	0,125
H_2O	22,375

The PCR started with an initial denaturation step at 94°C for 30 sec followed by 35 cycles with a denaturation step at 94°C for 30 sec, an annealing step at 55°C for 30 sec and an extension step at 68°C for 1 min.

3.3 Establishing an enzyme activity assay for FUT8

3.3.1 Preparation of the acceptor substrate for FUT8

3.3.1.1 Pronase digestion

3 g of fibrin were dissolved in 100 ml pronase buffer (100 ml 0,15 M Tris/HCl pH 7.8, 1 mM $CaCl_2$ + 0,02 % NaN_3) and denaturated for 25 min at 100 °C. The solution was cooled and 100 mg pronase (Roche, Vienna, Austria) dissolved in 1,5 ml pronase buffer were added. The digestion was incubated for 24 or 48 h at 37°C and then centrifuged for 15 min at 5000 rpm at 4°C. The pellet was discarded and the volume of the supernatant was reduced to approximately 2 ml in the rotavapor. This was then applied to a 120 x 15 cm Sephadex G25 fine (Amersham Bioscience, Freiburg, Germany) gel filtration column using 1% acetic acid in the liquid phase and collecting 2,2 ml per fraction. The fractions were measured for protein content at 280 nm. The released sugar were detected on a Alugram Sil G thin liquid chromatography plate with 0,2 mm silica gel (Machery-Nagel, Düren, Germany), which was previously coated with 100 µl orcinol solution (for 50 ml: 100 mg orcinol in 50 ml 20% H_2SO_4) and dried using a blow dryer. 1 µl per fraction were spotted on the plate, dried with a blow

dryer and incubated for 5 min at 100°C. The fractions which developed a reddish-brown colour were pooled and lyophilised.

3.3.1.2 Removal of the sialic acids

The lyophilised solution was dissolved in 50 ml 0,05 M H_2SO_4 and incubated for 1 h at 80°C. The volume was then reduced to approximately 2 ml in the rotavapor and applied to the same Sephadex G25 fine column as above. The protein and sugar content was analyzed as described for the pronase digestion. The fractions containing sugar were pooled and lyophilised.

3.3.1.3 N-glycopeptidase A (PNGase A) digestion

The lyophilized sample was dissolved in 1-2 ml PNGase A buffer (50 mM citrate/phosphate pH 5.0) and 6 µl PNGase A (N-glycopeptidase A from almonds prepared by Ing. T. Dalik, 5 mU/100 µl) were added. Incubation was done at 37°C for 24 or 48 h. The sample was then applied to the Sephadex G25 fine column and separation was performed as described before. The fractions containing sugar were pooled and the volume reduced to approximately 2 ml in the rotavapor. The pH was adjusted to 2 (checking with a pH stripe) with concentrated acetic acid.

3.3.1.4 Cation exchange chromatography

The sample was applied to a AG 50 wx2 column (100 ml column volume (CV), Bio-Rad, Vienna, Austria) in 2% acetic acid and then the column was washed with 1,5x column volumes of 2% acetic acid and further eluted with 0,4 M NH_4Ac pH 6.0 until reaching fraction 40. Fractions of 4 ml were collected. The acetic acid fractions containing the free sugars were pooled and lyophilised. The elution fractions glycopeptides and peptides were lyophilised and treated again with PNGase A.

3.3.1.5 ß-galactosidase digestion

The sample was dissolved in 1-2 ml of galactosidase buffer (50 mM natrium citrate pH 4.6, 0,04% NaN_3) and 1,5 µl galactosidase (ß-galactosidase from *Aspergillus oryzae* prepared by Ing. T. Dalik, 260 U/ml) was added. Incubation was done at 37°C over night. The sample was purified on the Sephadex G25 fine column and the fractions containing sugar were lyophilized. The acceptor was finally dissolved in 500 µl H_2O, quantified and frozen at -20°C.

3.3.1.6 Quantification

Quantification was carried out with the orcinol method. 50 µl of sample or standard were incubated with 200 µl orcinol reagent (for 200 ml dissolve 400 mg orcin in 25 ml H_2O; prepare 187.5 ml 60% H_2SO_4 (75 ml H_2O + 112,5 ml H_2SO_4) and cool down. Mix) for 45 min at 80°C and measured at 405 nm. A standard curve was obtained using glucose as the standard.

3.3.1.7 Quality control by fluorescence labelling and determination on HPLC

To approximately 30 nmol oligosaccharide 80 µl 2-aminopyrdine (2-AP) solution (stock solution: to 1 g 2-aminopyridin add 0,76 ml concentrate HCl in a glass tube. Store at -20°C. 2-AP solution: dilute the stock solution 1:3 with H_2O. Always prepare fresh solution) were added and incubated for 13 min at 100°C in the water bath. 4 µl $NaCNBH_3$ solution (weigh in10 mg $NaCNBH_3$ add 20 µl 2-aminopyridin solution and 30 µL H_2O. Prepare just before use) were added and incubated overnight at 90°C. The sample was applied to a 50 x 1 cm Sephadex G15 gel filtration column (Amersham Bioscience, Freiburg, Germany) in 1% acetic acid and 1 ml/fraction (total 30 fractions) were collected. 20 µl of sample were applied to the

Palpak Type S HPLC column (pore size 80 Å, particle size 5 µm, column size 250 x 4,5 mm, Takara, Otsu, Japan) with a flow rate of 1 ml/min, detection extinction/emission 310/380 nm and following gradient:

Time (min)	**0**	**5**	**45**	**46**	**47**	**53**
% buffer B	45	45	0	0	45	45

Buffer A: 50:50 acetonitril:3 % acetic acid in TEA buffer pH 7.3 (3% acetic acid adjusted with triethylamin to pH 7.3 containing 10% acetonitrile)

Buffer B: 100 % acetonitrile

3.3.2 Microsome preparation

For *A. lusitanicus* 1 ml cold isotonic buffer (5 mM imidazole/HCl pH 7.3 buffer with 250 mM sucrose) were added to approximately 20 mg connective tissue and homogenized with UltraTurrax at maximum speed 3 x 10 sec at 4°C. The sample was centrifuged at 6000 rpm for 25 min at 4°C. In the meantime ultracentrifugation tubes (1/2 x 2 inches, Beckman cat. #344057, Krefeld, Germany) were filled with 4,3 ml cold isotonic buffer and placed on ice. After centrifugation the supernatant was transferred to the ultracentrifugation tubes and placed in the pre chilled SW 55Ti rotor (Beckman, Krefeld, Germany). The sample was centrifuged at 4°C for 32'000 rpm for 1 h 15 min. The supernatant was discarded and the brown-orange pellet was resuspended in 200 µL 5 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 6.5 with 0.5 % Triton, transferred in an Eppendorf tube and homogenised by using a Potter-Elvehjem homogenisator. The solution was placed on ice and immediately used for an enzyme activity assay as described in chapter **Error! Reference source not found.**

For Bge cells the starting material was a confluent T75 flask cells which was harvested by centrifugation at room temperature for 5 min at 1500 x g. The cells were resuspended in 1 ml isotonic buffer and treated exactly as the connective tissue of *A. lusitanicus*. Enzyme activity was analyzed by radioactive enzyme activity assay (chapter 3.3.3).

3.3.3 Enzyme activity assays

For the radioactive enzyme activity assay, 20 µl of reaction contained 0.5 mM GnGn (for the structure see chapter 10) and 0.25 mM GDP-[^{14}C]-fucose (specific activity 5000 cpm/nmol/µl), 0.1 M MES pH 6.5, 0.1% Triton X-100, 10 mM $MnCl_2$ and 5 µl of the enzyme preparation. Incubation was carried of for 3 h at 37°C and was terminated by adding 500 µl 20 mM ice-cold sodium borate containing 2 mM EDTA. The samples were then applied to Pasteur pipettes filled with 500 µl AG1X8 (chloride form, 100-200 mesh, Bio-Rad, Vienna, Austria). Fucose and oligosaccharides were eluted with water, mixed with 4 ml Pico Aqua scintillation cocktail (Canberra Packard, Schwadorf, Austria) and counted. Each assay was run in triplicate; omitting the acceptor for the negative control. Radioactivity of the eluates obtained from the control was subtracted from that of the complete mixture for the calculation of enzyme activity (Staudacher et al, 1991).

In order to establish a HPLC-based assay for FUT8, the assay conditions were tested on a homogenised rabbit brain preparation where the activity has already been confirmed (Struppe & Staudacher, 2000). The assay conditions were as described below for the assay for FUT8. Incubation was carried out for 4 h at 37°C. The final reaction volume was 20 µl. The whole assay was labelled with 2-AP and analyzed by HPLC with a Palpak column as described in 3.3.1 for the acceptor.

For the enzyme activity assays of recombinant FUT8 following protocol was used:

Component	20 µl reaction
10 nmol acceptor substrate (chapter 3.3.1)	dried
0,4 M MES pH 7.0	5 µl
0,2 M $MnCl_2$	1 µl
4,25 mM GDP-fucose	2 µl
Enzyme	5 µl
H_2O	7 µl

The mixture was incubated for 3 h and overnight at 37 °C and then analysed either by HPLC or by matrix-assisted laser desorption/ionization - time-of-flight mass spectrometer (MALDI-TOF-MS) with an Autoflex Speed MALDI-TOF (Bruker Daltonics, Germany) equipped with a 1000 Hz Smartbeam.II laser in positive mode using 2% DHB in 50% acetonitrile as a matrix. Spectra were processed with the manufacturer´s software (Bruker Flexanalysis 3.3.80) (work done by Dr. Erika Staudacher).

3.4 Cloning and expression of a soluble incomplete form of *A. lusitanicus* FUT8 (amino acid residues 29 to 507)

3.4.1 Restriction of pVT-Bac

pVT-Bac (Tessier et al, 1991) was chosen as transfer vector for insect cells expression as it contains the honeybee melittin signal for enhanced protein secretion and an N-terminal 6x histidine tag for easier purification. A plasmid preparation of the vector (kind gift of Mag. Andreas Thader, Department of Chemistry, BOKU) was done as described in chapter 3.1.3.1 and restricted as follow:

Component	50 µl reaction
147 ng/µl pVT-Bac	3 µl
10x NEB Buffer 4	5 µl
20 U/µl *PstI HF* (New England Biolabs, Frankfurt am Main, Germany)	1 µl
20 U/µl *KpnI HF*(New England Biolabs, Frankfurt am Main, Germany)	1 µl
H_2O	40 µl

The reaction was incubated for 2,5 h at 37°C and purified by gel extraction with NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany) as described in chapter 3.1.3.2.

3.4.2 Restriction of pPICZα FlagHis1

pPICZα FlagHis1 (kind gift of Dr. Iain Wilson, Department of Chemistry, BOKU) is a modified version of the pPICα vector (Invitrogen, Vienna, Austria). It has following characteristics:

- AOX1 promoter for tightly regulated, methanol-induced expression of the gene of interest
- α-factor secretion signal for directing secreted expression of the recombinant protein
- N-terminal 6x histidine followed by flag tag for easier purification and detection

A plasmid preparation of the vector was done as described in chapter 3.1.3.1 and restricted as follow:

Component	**50 µl reaction**
1 µg/µl pPICZα FlagHis1	1 µl
10x NEB Buffer 4	5 µl
20 U/µl *PstI HF* (New England Biolabs, Frankfurt am Main, Germany)	1 µl
20 U/µl *KpnI HF*(New England Biolabs, Frankfurt am Main, Germany)	1 µl
H_2O	42 µl

The reaction was incubated for 2,5 h at 37°C and purified by gel extraction with NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany) as described in chapter 3.1.3.1.

3.4.3 Cloning FUT8 in pVT-Bac and pPICZα FlagHis1

First, it was necessary to amplify the known FUT8 sequence from the *A. lusitanicus* cDNA library (see chapter 3.2.2 for synthesis of the library). The PCR was done with Phusion High-Fidelity DNA Polymerase (New England Biolabs, Frankfurt am Main, Germany) as it has proofreading-activity using FUT8ArionINC_PstI_f (5' GTAGTActgcagGATGCAGCTAATCATGCAGAAC 3') and

FUT8ArionINC_KpnI(neu)_r (5' GATGATggtaccCTATTACCAATGGTTTCCAGC 3') as primers. Following protocol was used for amplification:

Component	**20 µl reaction**
5x Phusion HF Buffer	4 µl
10 mM dNTPs	0,4 µl
10 µM FUT8ArionINC_PstI_f	1 µl
10 µM FUT8ArionINC_KpnI(neu)_r	1 µl
1 ng/µl *A. lusitanicus* cDNA library	13,8µl
Phusion DNA Polymerase	0,2 µl

The PCR started with an initial denaturation step at 98°C for 30 sec followed by 40 cycles with a denaturation step at 98°C for 10 sec, an annealing step at 60°C for 30 sec and an extension step at 72°C for 1 min 40 sec and a final extension step at 72°C for 5 min.

The PCR product was purified by gel extraction with NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany) as described in chapter 3.1.3.2 and then restricted as following:

Component	**50 µl reaction**
19,6 ng/µl FUT8 PCR product	43 µl
10x NEB Buffer 4	5 µl
20 U/µl *PstI HF* (New England Biolabs, Frankfurt am Main, Germany)	1 µl
20 U/µl *KpnI HF*(New England Biolabs, Frankfurt am Main, Germany)	1 µl

The reaction was incubated for 2,5 h at 37°C and purified by NucleoSpin® Gel and PCR Clean-up (Machery-Nagel, Düren, Germany) as described in chapter 3.1.3.2.

FUT8 was then ligated to both vectors as following:

Component	10 µL reaction pVT-Bac	10 µL reaction pPICZα FlagHis1
84,4 ng/µl FUT8	0,5 µl	1 µl
13,1 ng/µl pVT-Bac	3,8 µl	-
200 ng/µl pPICZα FlagHis1	-	0,5 µl
10x T4 Ligase Buffer	1 µl	1 µl
400000 U/µl T4 DNA Ligase	1 µl	1 µl
H_2O	3,7 µl	6,5 µl

The reaction was incubated for 20 h at 16°C and 3,3 µl of it were used for heat-shock transformation of 50 µl of High Efficiency NEB 10-beta Competent *E. coli* cells (New England Biolabs, Frankfurt am Main, Germany). The transformed *E. coli* cells were plated on LB agar with 100 µg/ml Ampicillin in case of the pVT-Bac construct and on LB LS containing 25 µg/ml Zeocin (Invitrogen, Vienna, Austria) in case of the pPICZα FlagHis1 construct.

PCR screening was performed as described in chapter 3.1.4.2 using following primers:

Primer	Sequence
pVT BacHis Forward	5’ CATCTATGCGGATCCTATGCGG 3’
pVT BacHis Reverse	5’GCCGGACCAGTGAACAGAG 3’
5'AOX	5’ GACTGGTTCCAATTGACAAG 3’
3'AOX	5’ GCAAATGGCATTCTGACATCC 3’

Overnight cultures of the positive clones were grown and the plasmids were purified by mini prep with NucleoSpin Plasmid QuickPure (Machery-Nagel, Düren, Germany). This time sequencing was performed in both directions to assure that the sequence is correct. Therefore, pVT BacHis Forward/Reverse and 5'AOX/3'AOX respectively were used for sequencing by MWG Eurofins (Vienna, Austria).

3.4.4 Transformation of *P. pastoris* cells

For recombination of the plasmid into the *P. pastoris* cells genome it is necessary to linearize the plasmid before transformation. One of the plasmid preparations with the correct sequence was therefore restricted as following:

Component	50 µl reaction
1 µg/µl pPIC-FUT8 construct	25 µl
10x NEB Buffer 4	5 µl
10x Bovine serum albumin (BSA)	5 µl
20 U/µl *PmeI* (New England Biolabs, Frankfurt am Main, Germany)	1 µl
H_2O	14 µl

The reaction was incubated for 2,5 h at 37°C and 20 min at 65°C to inactivate the enzyme. Then, it was purified by NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany) as described in chapter 3.1.3.2.

The linearized and purified construct was used to transform electrocompetent X-33 *P. pastoris* cells (chapter 3.1.5.2). After 2 h of recovery on YPD the yeasts were plated on 100 µg/µl Zeocin YPD agar plates and incubated for 48 h at 30°C.

3.4.5 *P. pastoris* cells PCR screening

Before expression eight colonies from the transformation plate were picked and a masterplate was made. In addition a PCR screening was done using a modified version of the protocol described in (Lõoke et al, 2011), in short: the picked colonies were resuspended in 100 µL 200 mM lithium acetate with 1% SDS and incubated for 15 min at 70°C. Then 300 µl pure ethanol were added to it and centrifuged at room temperature for 5 min at maximum speed. The supernatant was discarded ant the pellet resuspended in 100 µl H_2O. Before PCR the solution was centrifuged for 1 min at 15000 g to get rid of the cell debris. 1 µl of the clear solution was used for PCR screening. In this case a PCR was made in parallel as a positive control to check if the genome extraction and the PCR reaction were done correctly. Therefore, an additional primer set, namely GAPDH_f (5' ATGACCGCCACTCAAAAGACC 3') and GAPDH_r (5' TTAGCAGCACCAGTGGAAGATG 3') was used.

All other parameters were exactly the same as following:

Component	30 µl reaction
10x Reaction Buffer	3 µl
10 mM dNTPs	0,6 µl
10 µM 5' AOX	0,3 µl
10 µM 3' AOX	0,3 µl
50 mM $MgCl_2$	0,9 µl
Template	1 µl
HybriPol DNA Polymerase	0,3 µl
H_2O	23,6

The PCR started with an initial denaturation step at 94°C for 3 min followed by 35 cycles with a denaturation step at 94°C for 30 sec, an annealing step at 55°C for 40 sec and an extension step at 72°C for 1 min 30 sec. The results were analyzed by gel electrophoresis on 1% agarose ethidium bromide gel.

3.4.6 Expression in *P. pastoris* cells

With each of the eight positive clones a 3 ml overnight culture was grown in YPD with 100 µg/µl Zeocin at 28°C. The day after, the overnight cultures were centrifuged and the pellet was resupended in 25 ml YPD. Pure methanol was added to a final concentration of 0,5 %. The cultures were incubated in a shaking incubator either at 28°C or at 16°C for 4 d and 5 d respectively. Each 24 h pure methanol was added to a final concentration of 0,5 %. After 4-5 d cells and supernatant were harvested by centrifugation and analyzed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) and Western Blot (WB).

3.4.7 *Sf*9 Cotransfection with Baculogold DNA, infection and scale up.

It was important to perform all the steps in fresh DNase/RNase free tube, so called "Sarstedt tubes" (Sarstedt, Wiener Neudorf, Austria).

2 µg of one of the plasmid preparations containing the correct sequence was mixed with 300 ng linearized BD BaculoGold™ Bright Baculovirus DNA (Becton Dickinson, Schwechat, Austria) and incubated for 5 min at room temperature. 1 ml insect cell medium without FCS was then added and incubated at room temperature for 30 min ("Tube A"). In the meantime 30 µl Lipofectin® Transfection Reagent (Invitrogen, Vienna, Austria) were mixed with 1 ml insect cell medium without FCS and incubated for 40 min at room temperature ("Tube B"). Tube A and B were then put together and incubated at room temperature for 30 min. In the meantime $9x10^5$ *Sf*9 cells were plated on a six-well plate and washed to time with insect cell

medium without FCS. The mix of tube A and B was then plated on the cells and incubated for 5 h at room temperature. After incubation, the mix was removed and 2,5 ml of insect cell medium with 3% FCS and 1x antibioticum/antimycoticum was put on the cells. After 5 d the supernatant was harvested and used to infect 2,5x10^6 *Sf*9 cells for 5 d. After a last scale-up on 7,5x10^6 cells, the supernatant was used to infect either *Sf*9, *Trichoplusia ni* High Five Cells or *Ascalapha odorata* Ao38 cells. 2-4 d after infection cells and supernatant were harvested by centrifugation and analyzed by SDS PAGE and WB.

3.4.8 SDS PAGE and Western Blot

200 µl of *P. pastoris* medium or cells were first methanol precipitated by adding 800 µl (4x volume) ice-cold methanol and incubating for 30 min at -80°C or overnight at -20°C. Insect cells were lysed by sonication, centrifuged for 5 min at 2000 x g at room temperature. Medium, soluble part and insoluble part after sonication were methanol precipitated as the *P. pastoris* sample. After centrifugation at 16000 x g for 40 min at 4°C, the supernatant was discarded, the pellets dried on air and resuspended in an appropriate volume of water. An equal volume of 2x sample buffer (31 mg DTT, 200 mg SDS, 5 ml 0.5 M Tris/HCl pH 6.8, 2,8 ml glycerin (87%), 2,7 ml H_2O, a few drops of bromphenolblue) was added to the samples and they were incubated at 96°C for 5 min before loading them on the 12.5% acrylamide gels.

The composition one gel was the following:

Component	**Separating gel 12,5% acrylamide**	**Stacking gel 4% acrylamide**
30 % (w/v) acrylamide	2499 µl	570 µl
1% (w/v) bisacrylamide	780 µl	390 µl
1.5 M Tris/HCl pH 8.8	1500 µl	
0.5 M Tris/HCl pH 8.8		750 µl
H_2O	1140 µl	1280 µl
10% SDS	60 µl	30 µl
10% Ammonium persulfate	36 µl	24
TEMED	3,6 µl	2,4 µl

The electrophoresis was performed in electrophoresis buffer (stock solution: 5x electrophoresis buffer 15 g/l Tris, 72 g/l glycin, 5 g/l SDS) using a Bio-Rad Mini-Protean II Cell (Bio-Rad, Vienna, Austria) at 200 V for 1h. The gels were either stained with Coomassie or electroblotted onto nitrocellulose (Laemmli, 1970). For Coomassie staining gels were first put into fixation solution (50% methanol, 7% acetic acid, 43% H_2O) for 30 min, then they were incubated with the Coomassie solution approximately 30 min (1% Coomassie brilliant blue R-250 in 40% methanol and 10% acetic acid) and the excess of staining solution removed by overnight incubation with 5% acetic acid. For electroblotting a sandwich of blotting paper-nitrocellulose-electrophoresis gel-blotting paper was first equilibrated in blotting buffer (10 % 250 mM Tris and 1.92 M glycine, 20% methanol, 70% H_2O). Then the sandwich was put in the blotting-machine (mini Protean II kit, Bio-rad, Vienna, Austria) and the transfer took place at 15 V for 40 min. The nitrocellulose sheets were then blocked with BSA 0.5% in TTBS buffer (100 mM Tris/HCl, pH 7.5, 100 mM NaCl, 0.01% Tween 20) and incubated either with Penta-His monoclonal antibody (Qiagen, Hilden, Germany; dilution 1:2000 in BSA 0.5% in TTBS) or, just for the *P. pastoris* sample, Monoclonal anti-FLAG M2 produced in mouse (Sigma-Aldrich, Vienna, Austria) for 1 h followed by three washing steps of 5 min in TTBS. The secondary antibody was for both alkaline phosphatase conjugated anti-mouse IgG from goat (Sigma-Aldrich, Vienna, Austria; dilution 1:2000 in BSA 0.5% in TTBS), which was incubated for 1 h. After washing twice for 5 min with TTBS and once for

10 min with distilled water, colour detection was performed using Fast BCIP/NPT (5-bromo-4-chloro-3-indolyl phosphate/nitrobluetetrazolium, Sigma-Aldrich, Vienna, Austria). The reaction was stopped with 5% acetic acid.

3.5 Cloning and expression of four soluble FUT8 fusion proteins

The first part of the missing 40 amino acids on the C-terminal of *A. lusitanicus* FUT8 belongs to the SH3 domain, which is highly conserved between FUT8s of different species. Even if its role is still not well understood, it may be important for the enzyme activity. Therefore, we built four different fusion proteins using the known soluble part of the *A. lusitanicus* FUT8 and replacing the whole SH3 domain with the SH3 domain of *Crassostrea gigas*, *Lottia gigantea*, *Caenorhabditis elegans* and *Drosophila melanogaster* respectively. The first two were the evolutionary nearest one to the slug but the function of their enzymes has not been proved until now. The last two are the evolutionary nearest organism where the activity of FUT8 was confirmed (Paschinger et al, 2005b).

3.5.1 Design of the constructs

All the fucosyltransferases sequences except the one of *Lottia gigantea* were obtained from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) with following accession numbers (in parenthesis): *Crassostrea gigas* (JH818267.1), *Caenorhabditis elegans* (AJ512486.1) and *Drosophila melanogaster* (AF441264.1). *Lottia gigantea*s sequence was obtained by blasting the known *A. lusitanicus* sequence on the *L. gigantea* project homepage http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html.The SH3 domains were identified by using the conserved domains tool from NCBI http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi.

The SH3 domain nucleotide sequences were synthesized as gBlocks gene fragments by Integrated DNA Technologies, Coralville, USA. In order to ligate them to the *A. lusitanicus* FUT8 sequence, it was necessary to add bp 1181 to 1427 from the *A. lusitanicus* sequence before the SH3 domain sequences and modify bp 1214 to 1219 with a silent mutation (ATCGCT → AAGCTT) in order to create an unique restriction site (*HindIII*). It was also necessary to introduce this silent mutation in the original sequence and therefore the following primer was designed: Fut8HindIII 5' GTACCTAAGCTTGCAGACTG 3'. To ligate the whole construct to pVT-Bac (chapter 3.4.1) a *KpnI* restriction site (GGTACC) plus six additional random nucleotides were added to the gBlocks after the stop codon.

3.5.2 Cloning of the constructs

- **Inserting the silent mutation in the *A. lusitanicus* FUT8 sequence:** In order to ligate the SH3 gBlocks to the *A. lusitanicus* sequence a silent mutation was created using a plasmid preparation with the pVT-Bac-FUT8 construct, a primer containing the mutation (see above) and following PCR protocol:

Component	20 µl reaction
5x Phusion HF Buffer	4 µl
10 mM dNTPs	0,4 µl
10 µM pVT BacHis Forward	1 µl
10 µM Fut8*HindIII*	1 µl
1 ng/µl pVT-Bac-FUT8	1,7 µl
Phusion DNA Polymerase	0,2 µl
H_2O	11,7 µl

The PCR started with an initial denaturation step at 98°C for 30 sec followed by 35 cycles with a denaturation step at 98°C for 10 sec, an annealing step at 55°C for 30 sec and an extension step at 72°C for 30 sec and a final extension step at 72°C for 5 min. The PCR product was purified by gel extraction with NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany) as previously described in chapter 3.1.3.2.

- **Restriction of the modified *A. lusitanicus* FUT8:** for ligation to pVT-Bac and the gBlocks, the modified FUT8 was digested with *PstI* HF and *HindIII* (both enzymes from New England Biolabs, Frankfurt am Main, Germany), as following:

Component	**50 µl reaction**
96 ng/µl modified *Arion* FUT8	10 µl
10x CutSmart Buffer	5 µl
20 U/µl *PstI HF*	1 µl
20 U/µl *HindIII*	1 µl
H_2O	33 µl

The mixture was incubated for 3 h at 37°C and purified by NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany) as previously described in chapter 3.1.3.2.

- **Restriction of the SH3 gBlocks:** the gBlocks as 200 ng lyophilized pellet. Before digestion they were resuspended in 20 µl TE buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA). For ligation to pVT-Bac and the modified FUT8, the gBlocks were digested with *KpnI HF* and *HindIII* (both enzymes from New England Biolabs, Frankfurt am Main, Germany), as following:

Component	**50 µl reaction**
10 ng/µl gBlocks	10 µl
10x NEB Buffer 2	5 µl
20 U/µl KpnI HF	1 µl
20 U/µl HindIII	1 µl
H2O	33 µl

The mixture was incubated for 3 h at 37°C and purified by NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany) as previously described in chapter 3.1.3.2.

- **Ligation of *A. lusitanicus* FUT8 to SH3 gBlocks:** sticky-end ligation of FUT8 to gBlocks was performed using T4 ligase following this protocol:

Component	**30 µl reaction**
6 ng/µl gBlocks	14 µl
18 ng/µl FUT8	8 µl
10x T4 ligase buffer	3 µl
400000 U/µl T4 ligase	1 µl
H2O	4 µl

The mixture was incubated overnight at 16°C and directly used for the next step.

- **Ligation of fusion protein to pVT-Bac:** the already digested pVT-Bac (chapter 3.4.1) was ligated to the fusion protein constructs as following:

Component	30 µl reaction
Fusion protein ligation	14 µl
14 ng/µl restricted pVT-Bac	1,5 µl
10x T4 ligase buffer	3 µl
400000 U/µl T4 ligase	3 µl
H2O	8,5 µl

The mixture was incubated overnight at 16°C and directly used for the next step.

- **Transformation in *E. coli* cells, cotransfection, infection scale-up, SDS-PAGE, WB and enzyme activity assay:** were all done exactly as described in chapter 0.

3.6 Cloning and expression of *Lottia gigantea* and *Crassostrea gigas* FUT8s

3.6.1 Design

The sequence of both fucosyltransferases (see previous chapter for sequence source) was analyzed using different transmembrane domain prediction softwares (Table 1):

Name	Method
HMMTOP	Hidden Markov Model
MEMSAT	Neural networks and SVMs
PHDhtm in PredictProtein	Multiple alignment-based neural network system
Phobius	Homology supported predictions
TMHMM	Hidden Markov Model
SVMTop2	Support Vector Machines

Table 1: Transmembrane prediction software.

None of these softwares (Table 1) predicted a transmembrane domain with a reasonable probability therefore the whole sequence of both FUT8s containing also the restriction sites for *PstI* and *KpnI* were synthesized by Integrated DNA Technologies, Coralville, USA.

3.6.2 Cloning

The genes were delivered in a vector as lyophilized 4 µg pellet which was resuspended in 10 µl H_2O (final concentration 400 ng/µl). In order to clone the genes in pVT-Bac this construct was first restricted with PstI HF and KpnI HF (both enzymes from New England Biolabs, Frankfurt am Main, Germany), as following:

Component	30 µl reaction
400 ng/µl FUT8 mini prep	3 µl
10x CutSmart Buffer	3 µl
20 U/µl *KpnI HF*	1 µl
20 U/µl *PstI HF*	1 µl
H_2O	20 µl

The mixture was incubated for 3 h at 37°C and purified by NucleoSpin Gel and PCR Clean-up (Machery-Nagel, Düren, Germany) as previously described (see chapter 3.1.3.2).

Transformation in *E. coli* cells, cotransfection, infection scale-up, SDS-PAGE, WB and enzyme activity assay: were all done exactly as described in chapter 0.

3.7 Purification of ß- galactosidase from *A. lusitanicus*

3.7.1 Sample preparation and ammonium sulphate precipitation

Five to seven frozen *A. lusitanicus* were washed carefully to remove as much slime as possible. The abdominal cavity was opened and the whole digestion system removed to avoid contamination with food. The remaining skin and viscera (approximately 12 g) were resuspended in 400 ml 50 mM Tris/HCl pH 7.5 and homogenised with Ultra Turrax for 30 sec at maximum speed. The homogenate was set to 40% ammonium sulphate and centrifuged for 45 min at 5000 rpm at 4°C. The precipitate was discarded and the supernatant slowly set to 80% of ammonium sulphate. After stirring at room temperature for 30 min the mixture was centrifuged for 30 min at 11000 rpm at 4°C. The supernatant was discarded, the pellet resuspended in 5 ml 10 mM sodium phosphate pH 7.0, 1,2 M ammonium sulphate and stirred overnight at room temperature.

3.7.2 Hydrophobic interaction chromatography

The column consisted of an octylsepharose CL4B gel CV of 50 ml. The column was washed with 2-3 CVs starting buffer (10 mM sodium phosphate pH 7.0, 1,2 M ammonium sulphate), then the sample was applied on the column and washed with 2 CV of starting buffer. Finally the gradient mixer containing 2 CV volumes of both, starting buffer and elution buffer (10 mM sodium phosphate pH 7.0), was applied to the column. Fractions of 4 ml drops were collected. Protein content and an enzyme activity (chapter 3.7.7) were measured. Fractions showing enzyme activity were pooled and the buffer changed to 50 mM Tris/HCl pH 8.2 in the ultrafiltration cell using a Millipore membrane with 10 kDA cut-off.

3.7.3 Anion exchange chromatography

Anion exchange chromatography was performed using a DE52 column with 50 ml CV. The column was washed with 2 CV starting buffer (50 mM Tris/HCl pH 8.2), then the sample was applied on the column and washed with 2 CV of starting buffer. Finally the gradient mixer containing 2 CV volumes of both, starting buffer and elution buffer (50 mM Tris/HCl pH 8.2, 1 M NaCl), was applied to the column. Fractions of 3,5 ml drops were collected. Protein content and an enzyme activity (chapter 3.7.7) were measured. Fractions showing enzyme activity were pooled and the volume was reduced to approximately 1 ml in the ultrafiltration cell.

3.7.4 Size exclusion chromatography

Size exclusion chromatography was performed using a Sephacryl S-200 column (120 cm x 1,5 cm) at 4 °C using 50 mM Tris/HCl pH 7.5. Fractions of 1,5 ml were collected and tested for enzyme activity of different exoglycosidases: ß-galactosidase, α-glucosidase, ß-xylosidase, α-fucosidase, α-mannosidase and ß-N-acetylglucosaminidase. Fractions showing high ß-galactosidase activity were pooled, buffer changed to 50 mM sodium citrate pH 4.6 and volume reduced to 200 µl.

3.7.5 Affinity chromatography

For the last step of purification first two different columns containing a substrate analogue for ß-galactosidase were tested using for both the same conditions:

- ρ-aminophenyl ß-D-thiogalactopyranoside (Sigma-Aldrich, Vienna, Austria), CV = 1 ml
- ρ-aminobenzyl 1-thio-ß-D-galactopyranoside (Sigma-Aldrich, Vienna, Austria), CV = 1 ml

The gels are suspended in 0,5 M NaCl and 0.02% thimerosal, therefore they were first washed with 3 CV 50 mM sodium citrate pH 4.6 (buffer 1). Then the sample was applied to the column, incubated for 15 min and 2 CV buffer 1 applied. 2 CV of each of the following buffers were applied sequentially:

- Buffer 2: 50 mM sodium citrate pH 4.6, 1 M NaCl.
- Buffer 3: 50 mM Tris/HCl pH 7.5.
- Buffer 4: 50 mM Tris/HCl pH 7.5, 1 M NaCl.

Fractions were collected and tested for enzyme activity of different exoglycosidases (see chapter 3.7.7): ß-galactosidase, α-glucosidase, ß-xylosidase, α-fucosidase, α-mannosidase and ß-N-acetylglucosaminidase.

3.7.6 SDS PAGE – silver staining

150 µl of each of the affinity chromatography fractions were methanol precipitated as previously described in chapter 3.4.8 and applied to the SDS-PAGE. Electrophoresis and Coomassie staining were also done as previously described in chapter 3.4.8. If no bands were clearly visible in Coomassie staining, silver staining was performed.

1	50% methanol, 12% trichloracetic acid + 2% $CuCl_2$	5 min
	washed thoroughly with H_2O	30 sec
2	10% methanol + 5% acetic acid	5 min
	washed thoroughly with H_2O	30 sec
3	0,01% $KMnO_4$	5 min
	washed thoroughly with H_2O	30 sec
4	10% methanol + 5% acetic acid	1 min
	washed thoroughly with H_2O	30 sec
5	10% EtOH	5 min
	washed thoroughly with H_2O	30 sec
6	H_2O	5 min
7	50 ml/gel freshly prepared 0,01% $AgNO_3$	5 min
	washed thoroughly with H_2O	30 sec
8	H_2O	1 min
9	50 ml/gel freshly prepared 1% K_2CO_3 + 0,1% formaldehyde	as needed
10	Stop staining with 5% acetic acid	-

3.7.7 Enzyme activity assay

For enzyme activity assay 25 µl sample were pipetted in a 96-well plate. As a substrate 25 µl 5 mM p-Nitrophenol-1-Hydroxy-4-nitrobenzol-sugar (pNP-sugars, Sigma-Aldrich, Vienna, Austria) in 0,1 M sodium citrate + 0,04% sodium azide pH 4.6 were added and the plate incubated for 2 h at 37°C. To stop the reaction 250 µl 0,4 M glycin/NaOH pH 10.4 were

added to each sample and the colour intensity resulting from hydrolysis measured on the plate reader at 405 nm. Following pNP-sugars were used:

- For β-galactosidase: pNP-β-D-galactopyranoside.
- For α-mannosidase: pNP-α-D-mannopyranoside.
- For α-glucosidase: pNP-α-D-glucopyranoside.
- For ß-xylosidase: pNP-β-D-xylopyranoside.
- For α-fucosidase: pNP-α-L-fucopyranoside.
- For ß-N-acetylglucosaminidase: pNP-N-acetyl-β-D-glucosaminpyranoside.

3.7.8 Mass spectrometry analysis of the purified ß-galactosidase

A clear single band on a Coomassie stained SDS-PAGE was excised, trypsin digested and analyzed by electrospray-ionisation mass spectrometry (work done by Dipl. Ing. Clemens Gruber). The peptides resulting from mass spectrometry analysis were used for homology search on different databases.

4 Results

None of the species we were working with (*A. lusitanicus*, *A. fulica*, *B. glabrata*) has been sequenced far enough while the project was ongoing. Some sequences were available from the genome project from *B. glabrata* but none of them was helpful for our aims. Therefore, our first strategy was to establish a full-length cDNA library in appropriate expression vectors, which are then transfected into the desired host cells (in our case *P. pastoris* and insect cells). The transfected cells containing the functional cDNA of the desired enzyme can be detected and isolated by the modified glycans expressed on the surface. This strategy was described by Larsen for the first time and successfully used since then in several cases (Ernst et al, 1989; Larsen et al, 1989; Prieto et al, 1997). For high- throughput screening FACS using the appropriate lectins for staining (*Lens culinaris* agglutinin in this case) was used.

The second more classical approach was to biochemically purify the enzyme as it has already been done in our lab for the alpha-1,3-fucosyltransferase from mung bean (Staudacher et al, 1995). Previous experiments showed no activity in the raw gastropod homogenate probably due to some disturbing agents. Therefore, in order to enrich for the little amounts of glycosyltransferases which are localized in the Golgi, microsome purification was the method of choice. Efficiency was proved with enzyme activity assays. Different enzyme activity assay approaches were tested for this scope.

Our last strategy was based on sequence homology by screening the full-length cDNA library. When aligning FUT8s from different species either using amino acids or nucleotide sequences a high degree of homology was found. Therefore, degenerate primers with a low degree of degeneracy were designed and used for a PCR. As glycosyltransferases are expressed in low amount, a nested PCR may help to get the correct PCR product. After having amplified and detected a piece of the sequence, the greatest challenge was to find the 5' and especially the 3' end of the sequence. The main method was using RACE but also southern blot, colony blot, inverse PCR and much more methods were tried in this work.

As none of the strategies worked completely, different constructs of FUT8s were also expressed: starting from a C-terminal missing version of the *A. lusitanicus* FUT8 to some fusion proteins containing the SH3 domain from related species and also the complete FUT8s from two other molluscs: *L. gigantea* (also a gastropod) and *C. gigas* (a bivalvia). Expression was mainly performed in insect cells and enzyme activity analyzed by MALDI-TOF.

Some work was also done on the classical biochemical purification of β-galactosidase from *A. lusitanicus* using a protocol which was well established in our lab. The plan was to obtain enough purified protein to get the peptide sequence and screen the cDNA library with degenerate primers.

4.1 Synthesis and screening of cDNA expression libraries from gastropods

For the first strategy poly(A) RNA was purified from the different gastropod species and full-length non-directional cDNA was synthesized. The cDNA was cloned in different vectors: first pSMART2if in order to characterize the library and then in three expression vectors (one for *P. pastoris* and two for insect cells). Some preliminary experiments were also done using mouse FUT8 and the insect cell expression system in order to prove the concept. Therefore, the insect cells were also preferred as expression system for the library. Through cell sorting using biotinylated *Lens culinaris* agglutinin and PE-streptavidin we tried to isolate transfected cells expressing the gastropod FUT8.

4.1.1 Double stranded cDNA synthesis and cloning

For the expression library ds cDNA was synthesized with the In-Fusion SMARTer cDNA Library Construction Kit. The starting material was approximately 100-200 ng of poly(A) RNA. In order to optimize the ds cDNA synthesis the appropriate number of cycles was tested (Figure 8)

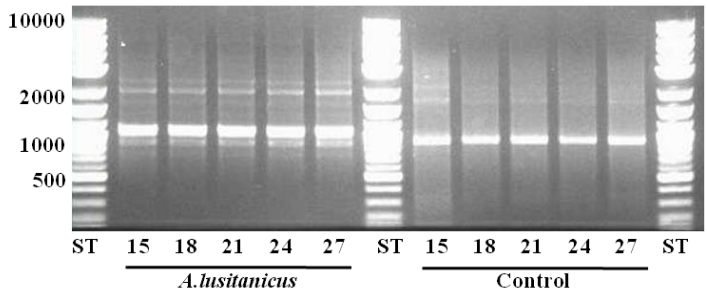


Figure 8: Ds cDNA synthesis from *A. lusitanicus*. ST: 2log ladder, 15-27: number of PCR cycles, Control: ds cDNA synthesis from human placenta RNA.

27 cycles was the optimal number of cycles in this case and the ds cDNA synthesized this way was cloned in the following vectors:

- pSMART2if: a cloning vector just for *E. coli* cells in order to characterize the library.
- pGAPZ B: expression vector for *P. pastoris* cells.
- pBakPAK8: is a well-established expression vector for insect cells, where recombination into the baculovirus genome is done *in vivo.*
- pENTR1a: is an expression vector for insect cells where recombination is done *in vitro*. It should be more efficient that the *in vivo* strategy and therefore improve the size and diversity of the library.

The constructs were transformed in *E. coli* cells and the *E. coli* cells plated on LB agar or LB LS agar in case of pGAPZ B with the appropriate antibiotic for selection. The libraries had following sizes:

- pSMART2if: $2x10^5$ clones.
- pGAPZ B: $5x10^4$ clones.
- pBakPAK8: 2×10^4 clones.
- pENTR1a : 7×10^2 clones.

PCR screening of all the libraries was performed to have a first overview of the diversity of the library, (Figure 9,Figure 10,Figure 11 and Figure 12).

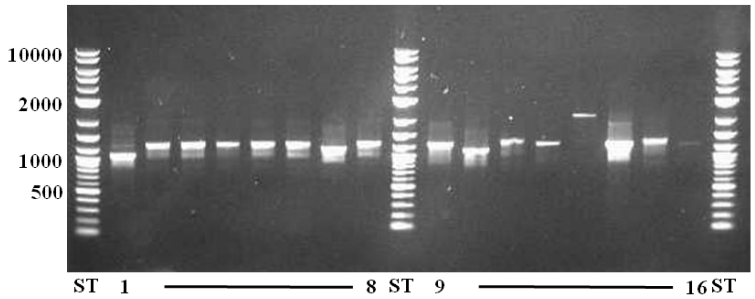


Figure 9: PCR screening of *A. lusitanicus* library in pSMART2if. ST: 2log ladder, 1-16: screened colonies; all colonies are positive and contain an insert between 800 and 1800 bp.

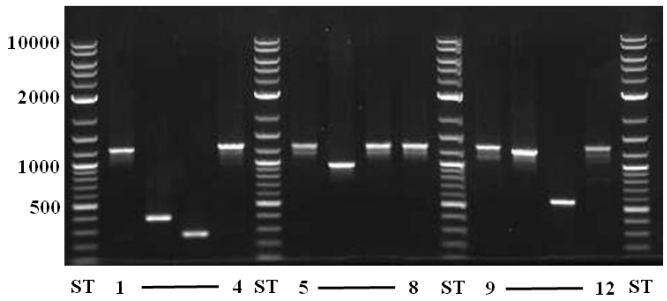


Figure 10: PCR screening of *A. lusitanicus* library in pGAPZ B. ST: 2log ladder, 1-12: screened colonies; colony 2 is an artefact and colony 3 is negative, all other colonies are positive and contain an insert between 300 and 1300 bp.

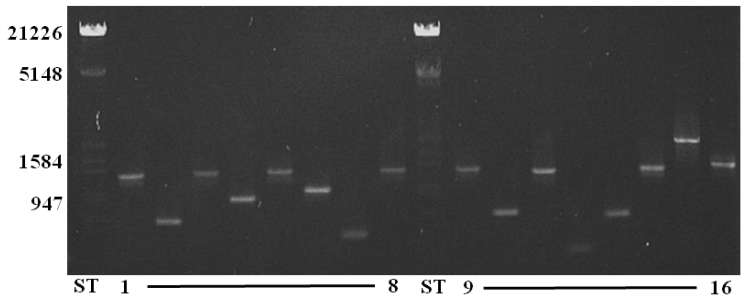


Figure 11: PCR screening of *A. lusitanicus* library in pBakPAK8. ST: λ DNA *HindIII/EcoRI* ladder, 1-16: screened colonies, colonies 5, 10 and 15 are negative, all other colonies are positive and contain an insert between 300 and 1700 bp.

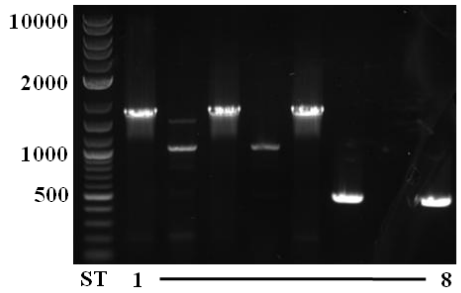


Figure 12: PCR screening of *A. lusitanicus* library in pENTR1a. ST: 2log ladder, 1-8: screened colonies; colonies 6 and 8 are negative, PCR of colony 7 did not work, all other colonies are positive and contain inserts of either 700 or 1000 bp.

Except for the pENTR1a library which was too small and too less diverse, a couple of clones from each library were sent for sequencing to confirm heterogeneity and lack of contaminations. The results were the following:

- The pSMARTif library consisted of 17% individual genes, 48% mitochondrial 16S rRNA gene, 30% unknown gene and 5% background (just vector).
- The pGAPZ B library consisted of 15 % individual genes, 45% mitochondrial 16S rRNA gene, 35% unknown gene and 5% background (just vector).
- The pBakPAK8 library consisted of 33% individual gene, 62% mitochondrial 16S rRNA gene and 5% background (just vector). Interestingly we found the entire ORF of a C-type lectin (Figure 13).

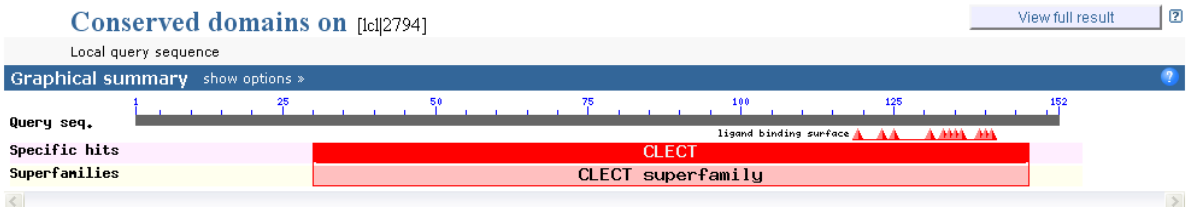


Figure 13: pBLAST output of one of the sequenced clones, showing the homology to a slug C-type lectin.

4.1.2 Construction of an expression library in *P. pastoris* cells

Electrocompetent X-33 *P.pastoris* cells were transformed with the library and after 48 h 1,5x 10^5 clones were counted (Figure 14A and B). After storage this library was no further used.

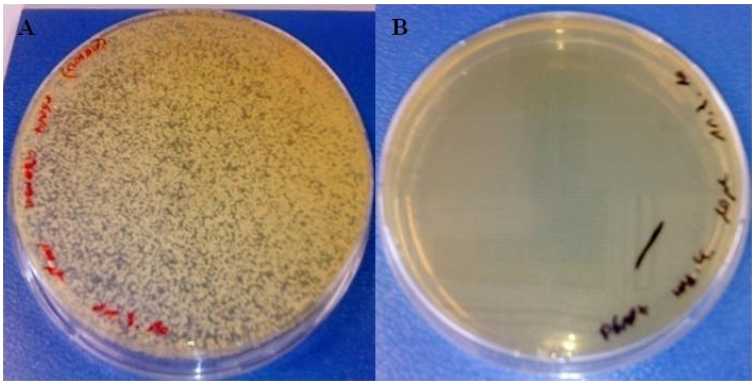


Figure 14A and B: **Transformed *P. pastoris*, A: with library, B: negative control without library.**

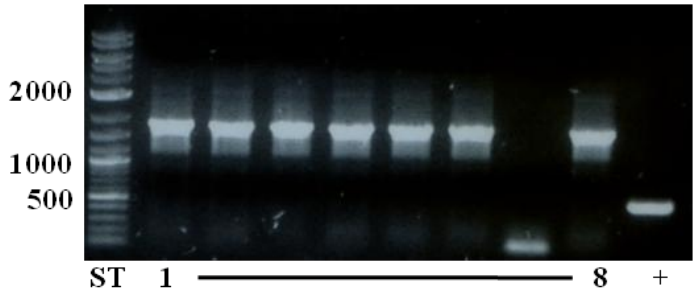
4.1.3 Construction of an expression library in *Sf*9 cells

As *Sf9* cells also express small amounts of FUT8 a proof of concept using a mouse FUT8 was made to show that it is possible to detect cells overexpressing recombinant FUT8.

4.1.3.1 Proof of concept with α-1,6-fucosyltransferase from *Mus musculus*

4.1.3.1.1 Cloning of mFUT8, recombination in the engineered baculovirus genome, creation of the baculovirus stock

Starting from a plasmid containing the entire sequence of mFUT8 a two step PCR was made in order to clone into pENTR1a vector. After ligation the construct was transformed in *E. coli* cells and PCR screening confirmed insertion of mFUT8 in pENTR1a (Figure 15).



.

Figure 15: PCR screening of colonies containing pENTR1a with the mFUT8 sequence. ST: 2 log ladder; 1-8: colonies screened; +: control (just vector); all but colony 7 are positive.

From all positive clones plasmid preparations were subjected to sequence analysis. Two of the clones had no mutations, one of these was used for the baculovirus.

4.1.3.1.2 FACS

$3*10^6$ *Sf9* cells were infected: either with the baculovirus expressing mFUT8 or with a baculovirus expressing a control protein.

Four days after infection the cells were stained with biotinylated *Lens culinaris* agglutinin as a primary antibody and streptavidin- PE as a secondary antibody. The stained cells were analysed by FACS and showed the expected result: the population with a high intensity signal was seven time as high in the mFUT8 expressing cells than in the control ones (Figure 16A and B)

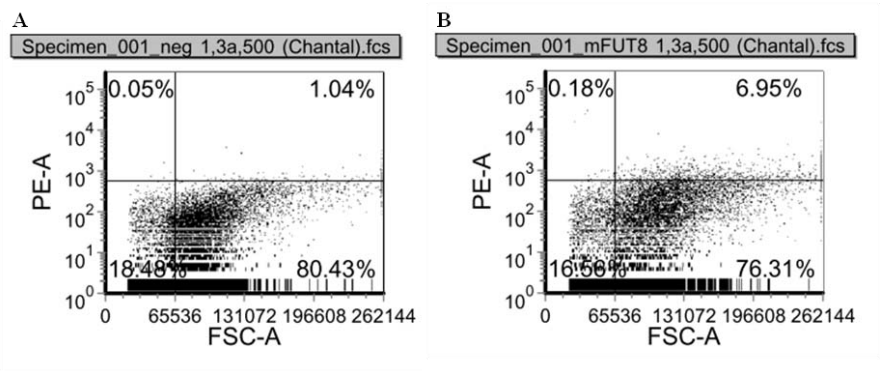


Figure 16A and B: FACS plot of cells expressing the control protein (A) and mFUT8 (B). 1.04% of the cells expressing the control protein show a high fluorescence signal, whereas in the cells expressing mFUT8 6,95% of the population show a high fluorescence signal.

4.1.3.1.3 Immunofluorescence microscopy

As an additional control a similar experiment was done using immunofluorescence microscopy. The intensity of the fluorescence signal of cells expressing mFUT8 was compared with the one of cells expressing a control protein by taking pictures of the stained cells with a confocal microscope. As seen in the FACS experiment, the cells expressing mFUT8 have a much more intense fluorescence than the control ones (Figure 17A and B and Figure 18A and B)

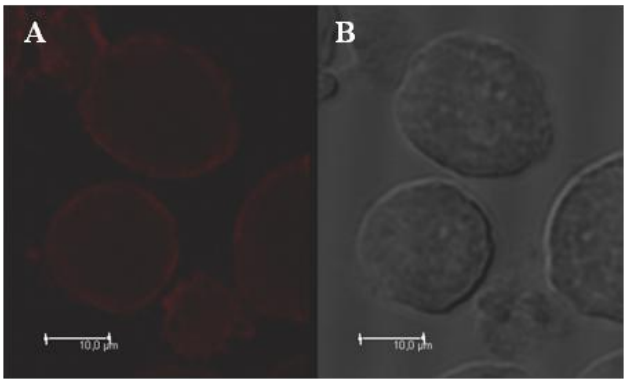


Figure 17A and B: Immunofluorescence (A) and brightfield (B) microscopy images of *Sf*9 cells expressing the control protein.

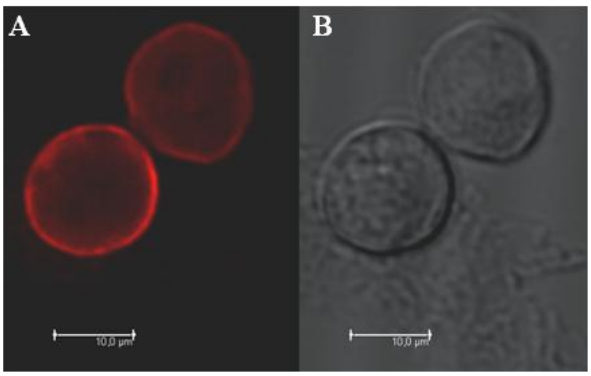


Figure 18A and B: Immunofluorescence (A) and brightfield (B) microscopy images of *Sf*9 cells expressing mFUT8.

4.1.3.2 Screening of the expression library in *Sf*9 by FACS

Before cell sorting preliminary FACS experiments were done to eventually adjust the staining protocol with both lectins specific for α-1,6-fucose (*Aleuria aurantia* lectin and *Lens culinaris* agglutinin). The two lectins gave a completly different result: when the library was stained with *Lens culinaris* agglutinin the population with a high fluorescence signal was higher than in the control and when using *Aleuria aurantia* lectin it was exactly the opposite way (Figure 19A and B and Figure 20A and B).

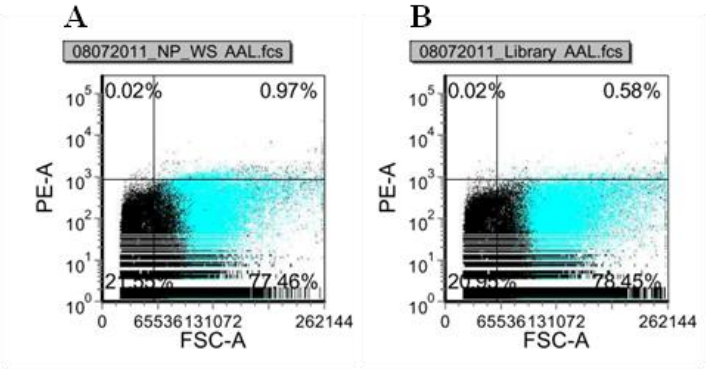


Figure 19A and B: FACS plot of the *Aleuria aurantia* lectin staining. In (A) the control cells have 0,97% of the population with a high signal, whereas in (B) the library cells 0,57% of the population has a high signal.

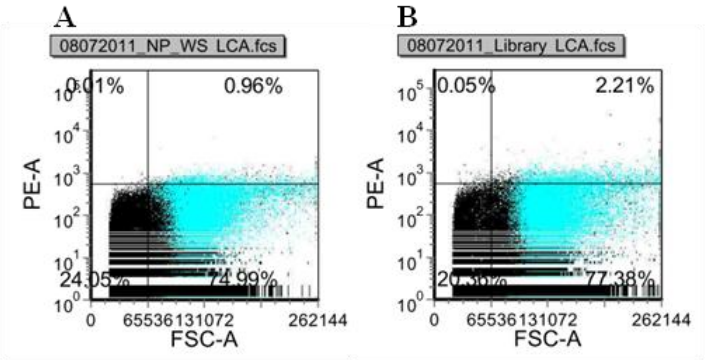


Figure 20A and B: FACS plot of the *Lens culinaris* agglutinin staining. In (A) the control cells have 0,96% of the population with a high signal, whereas in (B) the library cells 2,21% of the population has a high signal.

Lens culinaris agglutinin and the staining protocol used for the staining with it were used for the further FACS experiments.

4.1.3.3 Cell sorting

5% of the cells with the highest fluorescence intensity were sorted and the cells suspension used for infection of $7*10^5$ *Sf*9 cells and for further virus propagation. In parallel cells were infected with the virus containing the original, unsorted library as a control. Before going for the next round of sorting, the infection was analyzed by FACS to confirm the efficiency of sorting, as the population with a high signal should increase. The whole procedure was repeated three times for a total of three rounds of sorting. The following images show the analytical FACS of the first two rounds, where it seems that there is an enrichment of the library (Figure 21A and B and Figure 22A and B).

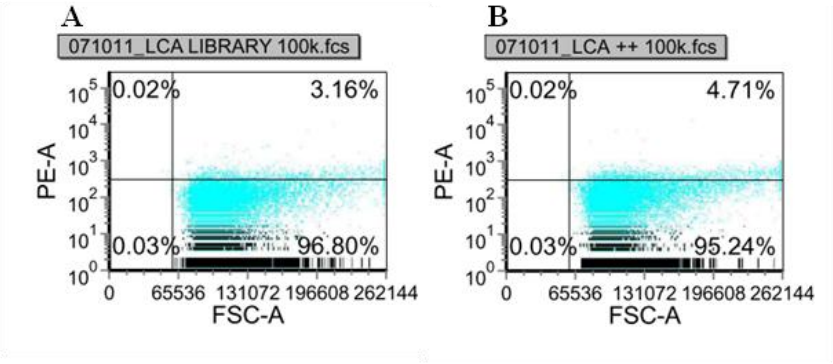


Figure 21A and B: FACS plot of the first round of sorting. It shows in (A) the library expressing cells with 3,16% of the population having a high signal and in (B) the cells from the first sorting round with 4,71% of the population.

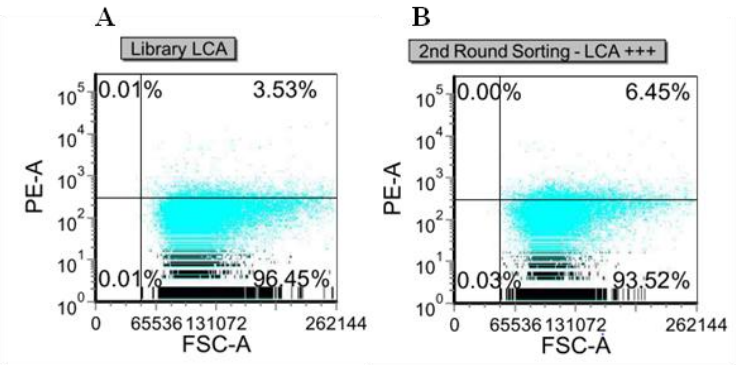


Figure 22A and B: FACS plot of the second round of sorting. It shows in (A) the library expressing cells with 3,53% of the population having a high signal and in (B) the cells from the second sorting round with 6,45% of the population.

The percentage of population with high fluorescence signal in the original, unsorted libraries in the first (3,16%) and second round (3,53%) was different as they were two different infections and some variability was expected. Nevertheless, the difference between the unsorted and sorted libraries was still two little after two rounds of sorting. To check if there was a constant increase of the population with a high signal or if it was just a statistical fluctuation, after the third round, cells were infected in parallel with the unsorted library and round one to three of sorting. As an additional control a PCR of the three baculovirus stocks containing the sorted libraries was done.

4.1.3.4 FACS control of sorting after three rounds

A constant enhancement of the population showing high fluorescence intensity should be visible if there was a real enrichment of the library for the target gene. The geometric mean and the percentage of cells with a high fluorescence should constantly increase from round to round. This was tested by infecting cells with the original library and the virus of the three sorting rounds. As already proven by the PCR there is no real enrichment of the library (Figure 23).

	Filename	% of gated cells	Geometric Mean
Unsorted cells	021211_LCA Library.fcs	4.31	514.27
1st round	021211_LCA++.fcs	8.24	489.85
2nd round	021211_LCA+++.fcs	5.77	482.35
3rd round	021211_LCA 4+_1.fcs	6.62	510.23
	021211_LCA 4+_2.fcs	6.26	482.59

Figure 23: Summary of the FACS control of the three rounds of sorting. The geometric mean and the percentage of cells show that there is no real enrichment of the library.

4.1.3.5 PCR control of sorting after three rounds

In theory, during sorting there should be a reduction in gene diversity of the library and enrichment for the target gene, which should be visible if a PCR of the baculovirus stocks containing the sorted libraries. The control PCR showed a lack of diversity and no enrichment for the target gene as the lanes look all very similar (Figure 24).

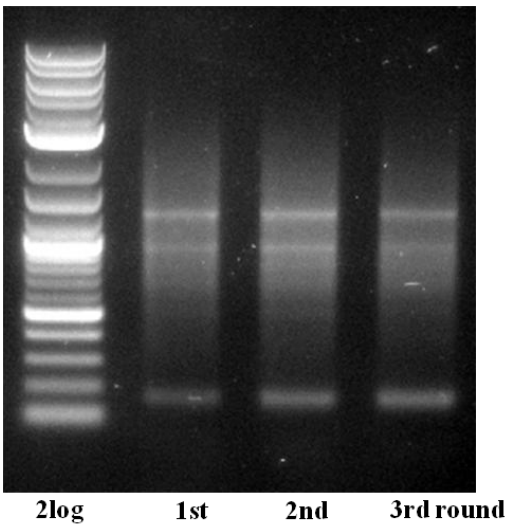


Figure 24: PCR of the baculovirus stocks after the three rounds of sorting showing no difference. 2log: 2log ladder, 1st-3rd round: round of sorting.

4.2 Identification of FUT8 by PCR in directional cDNA libraries from gastropods

This strategy was very similar to the first strategy in its starting steps. It started again with poly(A) purification from gastropod and ds cDNA synthesis. All cDNAs were oriented correctly, which enhanced the chances to find the correct sequence. Instead of cloning the library, degenerate primers were designed by aligning amino acid and nucleotide sequences of FUT8s from different species. The most successful approach was aligning the nucleotide sequences of human, mice, fruit fly and *C. elegans* FUT8s and by nested PCR.

To find the 5' and 3' end of the sequence mainly RACE-PCR was used. As 3' RACE did not work, other strategies were used such as southern blot, colony blot and inverted PCR.

4.2.1 Alignment of different FUT8 sequences and selection of primers for homology search

The alignment which led to a successful hit was the alignment of the nucleotide sequences of *Homo sapiens* (CAA76986.1), *Mus musculus* (AAH10666.1), *Drosophila melanogaster* (AAF48079.1) and *Caenorhabditis elegans* (CAD54736.1). A short piece of the aligned sequences and the selected primer are shown here (Figure 25)

```
Human            CTAGTGTGTACTTTTTCATCCCAGGTCTGTCGAGTTGCTTATGAAATTATGCAAACACTA 1059
Mouse            CTAGTGTGTACTTTTTCATCCCAGGTCTGTCGGGTTGCTTATGAAATCATGCAAACCCTG 1446
Drosophila       CTGGTGTGCACCTTCTCGTCGCAGGTGTGTCGCGTGGCCTACGAGATAATGCAGACGATG 1584
Caenorhabditis   TTGGTTTGCACTTTTTCAAGTCAGGTATGCCGAATGGGCTACGAGTTACGACAACCATCA 1407
                  * ** ** ** ** **     **** ** **  * *  ** **  *    **  *
Human            CATCCTGATGC--CTCTGCAAACTTCCATTCTTTAGATGACATCTACTATTTTGGGGGCC 1117
Mouse            CATCCTGATGC--CTCTGCGAACTTCCATTCTTTGGATGACATCTACTATTTTGGAGGCC 1504
Drosophila       TATCCGGATGCAGCGCATCGG--TTCAAGTCGCTGGACGACATATACTACTACGGTGGCC 1642
Caenorhabditis   GGAGCTGACGA--TGGATCAAAGTTCCATTCACTTGACGATATATACTATTTTGGAGGTC 1465
                     * ** *        *    *** * **  * ** ** ** ***** *  ** ** *
Human            AGAATGCCCACAATCAAATTGCCATTTATGCTCACCAACCCCGAACTGCAGATGAAATTC 1177
Mouse            AAAATGCCCACAATCAGATTGCTGTTTATCCTCACAAACCTCGAACTGAAGGAAATTC 1564
Drosophila       AGAATGCGCACAATCGCCGCGTCGTTATCGCCCACAAGCCGCGCACGCACGAGGATCTAC 1702
Caenorhabditis   AACAAGCTCATGAAGTAATTGTTATCGAAGATCATATTGCTCAGAACAATAAAGAAATCG 1525
                 *  * ** **  *       *   *       **     * *  *       * **  *
Human            CCATGGAACCTGGAGATATCATTGGTGTGGCTGGAAATCATTGGGATGGCTATTCTAAAG 1237
Mouse            CAATGGAACCTGGAGATATCATTGGTGTGGCTGGAAACCATTGGGATGGTTATTCTAAAG 1624
Drosophila       AGCTGCGTGTTGGTGACCTCGTCTCGGTGGCTGGTAACCATTGGGATGGCAATTCCAAGG 1762
Caenorhabditis   ATTTGAAAGTTGGTGATAAAGTTGGAATTGCTGGAAATCATTGGATGGATATTCTAAAG 1585
                    **     *** **     *     * ***** ** ****** *  ***  **** ** *
```

Figure 25: alignment of the nucleotide sequences of FUT8s from *Homo sapiens* (CAA76986.1), *Mus musculus* (AAH10666.1), *Drosophila melanogaster* (AAF48079.1) and *Caenorhabditis elegans* (CAD54736.1). In red the outer set of primer and in green the inner set of primer for nested PCR are marked.

The domains involved are shown in the alignment of the amino acid sequences. These domains are typical for the FUT8 family and are highly conserved (Figure 26).

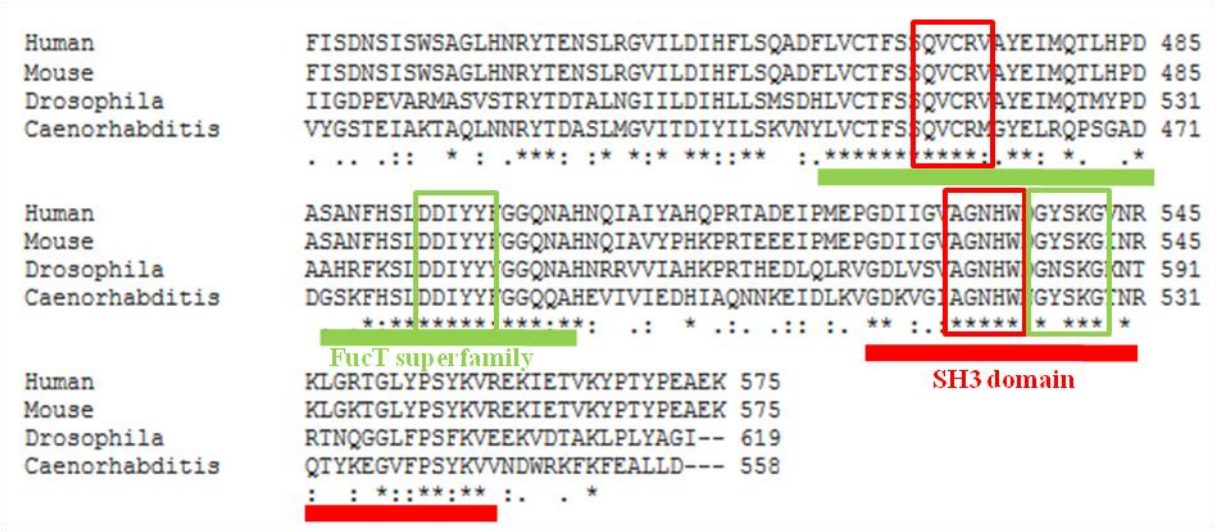


Figure 26: alignment of the amino acid sequences of FUT8s from *Homo sapiens* (CAA76986.1), *Mus musculus* (AAH10666.1), *Drosophila melanogaster* (AAF48079.1) and *Caenorhabditis elegans* (CAD54736.1). In red the outer set of primer and in green the inner set of primer for nested PCR are marked. The FucT superfamily conserved sequences and the SH3 domain are also shown.

The sequence of the primers was the following (Table 2):

Primer name	Sequence
ntFUT8_1f	5' CAGGTVTGYCGVRTKG 3'
ntFUT8_2f	5'GAYGAYATMTACTAYT 3'
ntFUT8_2r	5' YCCAATGRTTWCCAGC 3'
ntFUT8_1r	5' CYTTRGAATWDCCATYC 3'

Table 2: Primer sequences for the nested PCR. ntFUT8_1f and ntFUT8_1r were used for the first PCR and ntFUT8_2f and ntFUT8_2r were used for the second PCR.

4.2.2 cDNA Library construction

In order to provide a more diverse pool of genes, new cDNA libraries were generated for all snails. Here, just cDNA library from *A. lusitanicus* is shown (Figure 27). As done for the other libraries, the number of optimal cycles for ds cDNA synthesis was determined (Figure 27).

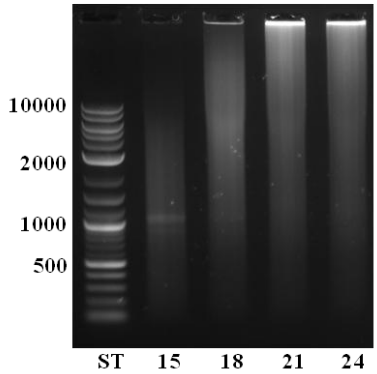


Figure 27: Ds cDNA synthesis from *A. lusitanicus*. ST: 2log ladder, 15-24: number of PCR cycles.

15 was the optimal number of cycles. The ds cDNA was not further processed and stored at -20°C.

4.2.3 PCR strategy to identify FUT8

4.2.3.1 Nested PCR

Using the protocol described in chapter 3.2.3.1 nested PCR was performed. Using the primer set listed in the table above, bands of the correct size were detected after the second PCR in each library (Figure 28).

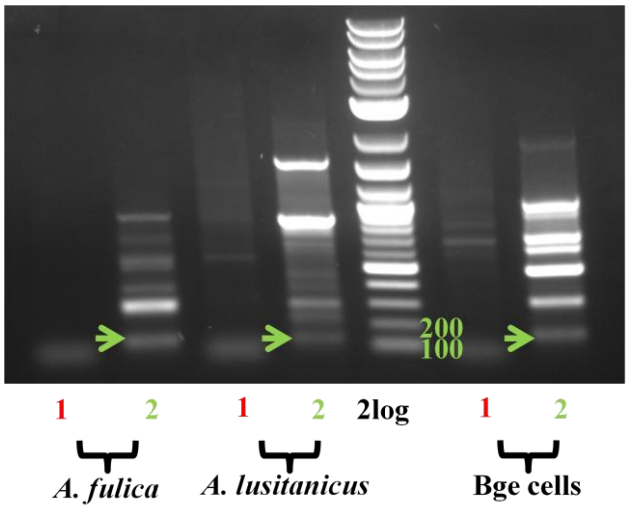


Figure 28: Gel of the nested PCR. The green arrows show the bands with the correct size. 1: first PCR, 2: second PCR, 2log: 2log ladder.

4.2.3.2 T/A- cloning and sequencing of the nested PCR fragments

All three inserts of approximately 120 bp were cloned through T/A cloning in the pGEM-T vector and transformed in *E. coli*. There were no clones on the *B. glabrata* plate, therefore PCR screening was performed just for *A. fulica* and *A. lusitanicus* plate (Figure 29).

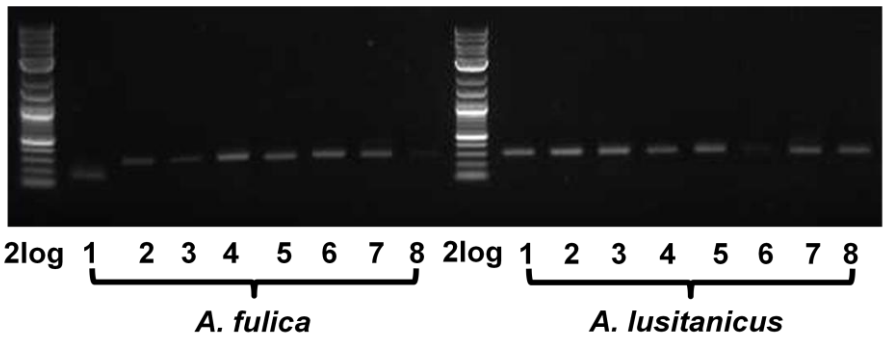


Figure 29: PCR screening of the *A. fulica* and *A. lusitanicus* plate after cloning the insert of the nested PCR. 2log: 2log ladder, 1-8: clones tested.

In *A. fulica* clones 2-8 were positive and in *A. lusitanicus* all clones but 6 were positive. For each gastropod four clones were subjected to sequence analysis. None of the clones from *A. fulica* gave interesting hit on pBLAST whereas clone 1 and 2 from *A. lusitanicus* where homologous to *C. gigas* FUT8 (Figure 30).

```
>gb|EKC39396.1|  Alpha-(1,6)-fucosyltransferase [Crassostrea gigas]
Length=575

 Score = 60.1 bits (144),  Expect = 2e-09, Method: Composition-based stats.
 Identities = 24/43 (56%), Positives = 29/43 (67%), Gaps = 0/43 (0%)

Query  1    DDIYYYGGQNAHDLSVIEAHHGNTENMLDILPRDTVGIAGNHW  43
            DDIYY+GGQN H +  +E H    E  +D+ P D VGIAGNHW
Sbjct  492  DDIYYFGGQNGHSVEAVEKHVKQNEKEIDLEPGDLVGIAGNHW  534
```

Figure 30: pBLAST output after translating the nucleotide sequence, showing homology to *C. gigas* FUT8.

4.2.4 Strategies to find the whole FUT8 sequence

4.2.4.1 5'RACE

After synthesis of 5'RACE ready cDNA, two specific primers based on the known sequence were designed. Again, one was an outer and the other an inner primer in order to make a nested PCR. When using just one PCR, no PCR product was visible on the gel, but after the second PCR a band in the correct size at 1600 bp appeared (Figure 31).

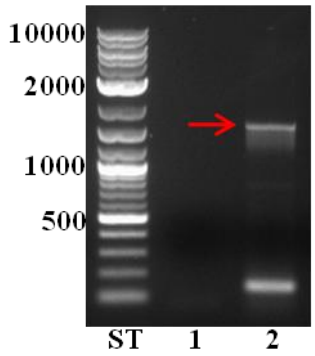


Figure 31: Successful nested 5'RACE PCR. The fragment of the correct length is highlighted with the red arrow. ST: 2log ladder **,** 1: first PCR, 2: second PCR.

4.2.4.2 Blunt end cloning and sequencing of the 5'RACE fragment

The insert of approximately 1600 bp was cloned by blunt-end cloning in the pUC19 vector and transformed in *E. coli* cells. PCR screening was performed and the positive clones sent for sequencing. The translated sequence showed a high homology to different FUT8 and had apparently all domains that are characteristic for the enzyme including the GDP- fucose binding sites and the beginning of the SH3 domain (Figure 32 and Figure 33).

```
>FUT8_1_536
MKQWKVIVLLLSFWLCIVLYMTNSVPSGMDAANHAERSLSRAMEELDKLYQQNKKLEQLIMQLKKNETFPHGLRD
NVTVEKLEFRLLRASNELSQIADSTQTKLTYEGEQARRKAENTVKELWYFLNSQLKKLDHIDNNDNHLTERISKL
KKDLEGYRRTTLEDFEKLRKTNHADDYRLQKSQELGDLVQRRLEYIQNPVNCKTAKKIVCNLHKGCGFGCQLHHI
TYCLIAAYAMERTLILDSKGWRYSPTGWESVFEPLSKTCSQVNNEGSRTHWRSSVEEMSSYDIIDLPIVDSLHPR
PGFMPLSVPADLAHDISIFHGDPSVWWIGHIVRYLFRLRPMVLQDVVNAGKKMGFQNTIVGVHVRRTDKIDLEAA
FHPLLEYMLHVGEYFDQLERTQSNITRRVYLASDDPNVLTEAQKQYPAYMFISDRSISQSASLGTRYTDNSLRGI
VIDIYYLARCDFLVCTFSSQVCRVAYEIMQTLHGDASKNFRSLDDIFYYGGQNAHDLSVIEAHHGNTENMLDILP
GDAVGIAGNHW
```

Figure 32: Amino acid sequence of *A. lusitanicus* FUT8.

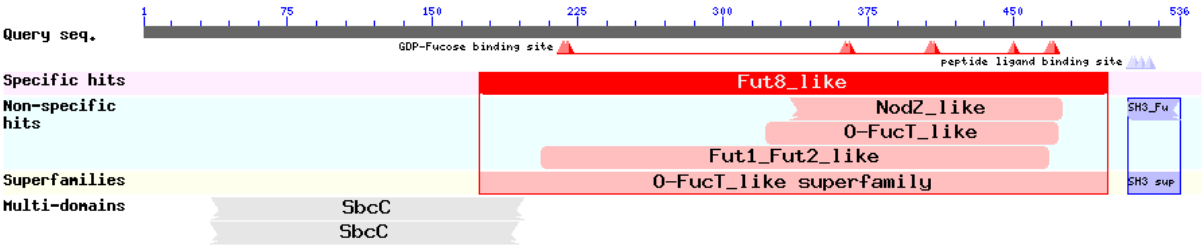


Figure 33: Domains recognized by the cds database at http://blast.ncbi.nlm.nih.gov/Blast.cgi . All typical FUT8 domains are present.

The translated sequence was also put into a transmembrane prediction software http://www.cbs.dtu.dk/services/TMHMM/ and it showed the classical type two Golgi membrane protein structure with a short N-tail, a transmembrane domain and the soluble catalytic domain on the C-terminus (Figure 34).

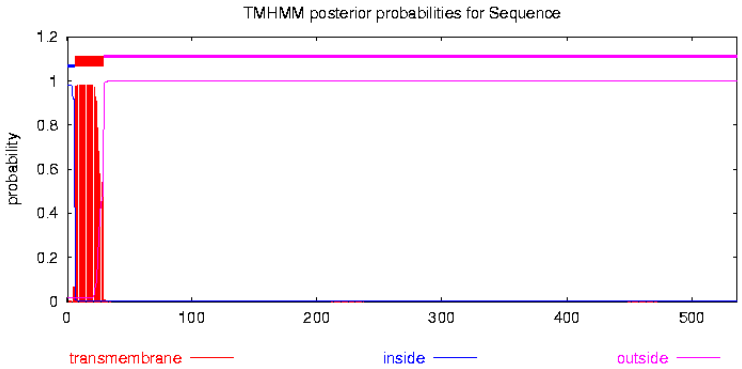


Figure 34: Plot of the transmembrane prediction software showing the classical type two Golgi membrane protein structure.

4.2.4.3 3'RACE

After synthesis of 3'RACE ready cDNA several primers (listed in 3.2.4.4) were tried in order to find the approximately 130 bp missing. Also the following possibilities were tried for 3' RACE:

- Different primer combinations as outer and inner primer for nested PCR.
- Different annealing temperatures, extension times, and cycle number.
- Variations of the 5' RACE protocol described in chapter 3.2.4.2 for Hybripol Polymerase and OneTaq Polymerase.

None of these strategies were successful even if often fragments of the correct size were found. The fragments were often some non-sense sequences or FUT8 splicing variants stopping for example in the middle of the sequence (Figure 35).

```
1100#5_M13uni-21      CATCACATCACCTACTGTCTTATAGCAGCATATGCCATGGAAAGAACTCTGATACTTGAT 60
fut8                  CATCACATCACCTACTGTCTTATAGCAGCATATGCCATGGAAAGAACTCTGATACTTGAT 60
                      ************************************************************

1100#5_M13uni-21      TCCAAAGGCTGGAGGTACTCACCCACAGGCTGGGAGTCTGTGTTCGAACCTCTTAGCAAG 120
fut8                  TCCAAAGGCTGGAGGTACTCACCCACAGGCTGGGAGTCTGTGTTCGAACCTCTTAGCAAG 120
                      ************************************************************

1100#5_M13uni-21      ACTTGTTCACAAGTAAACAATGAAGGTTCCAGAACCCACTGGAGATCATCTGTGGAAGAA 180
fut8                  ACTTGTTCACAAGTAAACAATGAAGGTTCCAGAACCCACTGGAGATCATCTGTGGAAGAA 180
                      ************************************************************

1100#5_M13uni-21      ATGAGCTCGTATGACATCATTGACCTTCCTATCGTTGACAGCCTTCATCCTCGTCCGGGT 240
fut8                  ATGAGCTCGTATGACATCATTGACCTTCCTATCGTTGACAGCCTTCATCCTCGTCCGGGT 240
                      ************************************************************

1100#5_M13uni-21      TTTATGCCACTCTCAGTGCCCGCAGATCTTGCTCATGACATCTCAATATTTCATGGTGAC 300
fut8                  TTTATGCCACTCTCAGTGCCCGCAGATCTTGCTCATGACATCTCAATATTTCATGGTGAC 300
                      ************************************************************

1100#5_M13uni-21      CCGTCGGTGTGGTGGATAGGTCACATTGTTCGTTACTTGTTCAGATTAAGACCAATGGTA 360
fut8                  CCGTCGGTGTGGTGGATAGGTCACATTGTTCGTTACTTGTTCAGATTAAGACCAATGGTA 360
                      ************************************************************

1100#5_M13uni-21      CTTCAGGATGTGGTCAATGCTGGCAAGAAGATGGGATTCCAAAACACCATTGTTGGGTAA 420
fut8                  CTTCAGGATGTGGTCAATGCTGGCAAGAAGATGGGATTCCAAAACACCATTGTTGG---- 416
                      ********************************************************...

1100#5_M13uni-21      GTTAGATGTGGGAGATAGACAAAGACATGTATCAGTTATATGAGGACTATCATCAT--AG 478
fut8                  GGTACATGTGCGTAGAACAGATAAAATTG-ATTTGGAAGCAGCCTTCCATCCCCTTTTGG 475
                      *.**.***** .*:..:*.*.*:*.*.:**.**.:*.:*.:*.:*.  .:*.***. .*:*..*
```

Figure 35: Alignment of the splicing variant (1100'5_M13uni -21) to the known FUT8 sequence (fut8). In yellow the putative splicing signal.

4.2.4.4 Colony blot

As an additional strategy to find the whole sequence, a colony blot on the cloned cDNA library was made. After transferring the colonies on a positively charged nylon membrane, the cells were lysed, the (plasmid) DNA was fixed by baking and a DIG labelled probe (fragment of the known *A. lusitanicus* FUT8 sequence) was hybridized to it. After incubation with an anti-DIG alkaline phosphatase conjugate and the following colour reaction (Figure 36), the putative positive colonies were picked from the original plate and plasmid preparations done. Sequencing of the plasmid showed that the binding of the probe was unspecific and no positive hit was found.

Figure 36: Section of the colony blot membrane after the colouring reaction showing one positive colony.

4.2.4.5 Southern blot

As many splicing variants were found when doing 3'RACE and as alternative method to find the 3' end a Southern Blot was performed. Therefore, the directional cDNA synthesized as described in chapter 3.2.2 was cut near its end with *MslI* and the probe was designed to bind near to the 3' end. If a signal could be detected at approximately 300 bp, a gel could be run under the same conditions as the southern blot gel and a band excised at the same height.

A positive control was made using the known *A. lusitanicus* FUT8 sequence. The protocol is very similar to the colony blot one using a DIG probe and anti- DIG alkaline phosphatase conjugate. This time detection was done by chemiluminescence. Even after 30 min of exposition no signal was visible in the library. Just on the positive controls the bands were very bright (Figure 37A, B and C).

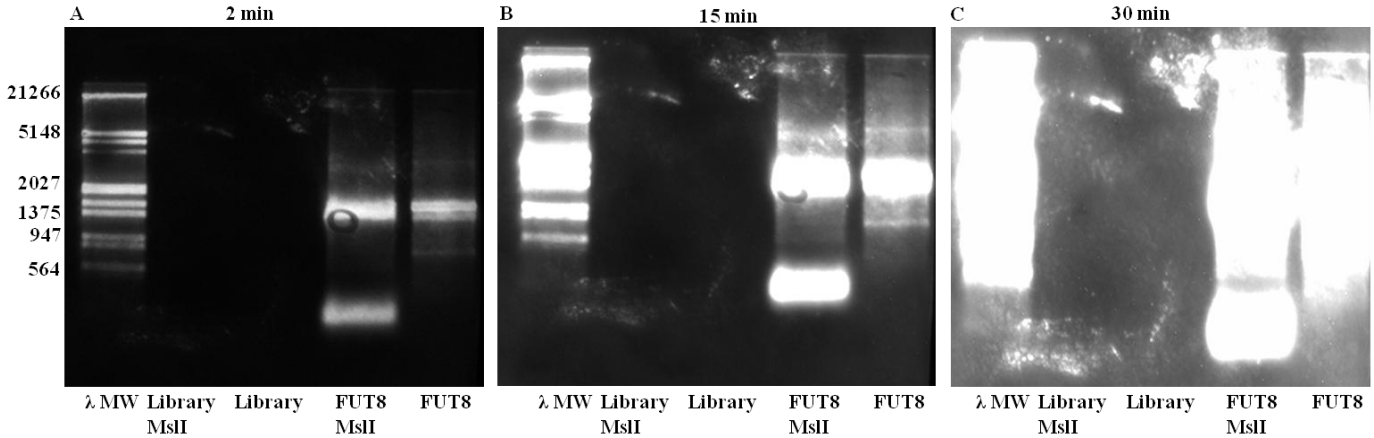


Figure 37A, B and C: Southern blot with chemiluminescence detection after 2 (A), 15 (B) and 30 (C) min exposition. The positive controls (FUT8 *MslI* and FUT8) show bright bands of the correct size but no bands are visible in the libraries. λ MW: Lambda *HindIII/EcoRI* ladder.

4.2.4.6 Inverse PCR on blunt end self ligated and sticky end self ligated cDNA

The idea behind this experiment was that during the ligation reaction the cDNA molecules rather ligate with themselves than with others. Therefore, it is possible to use primers on the known sequence to do an inverse PCR and amplify the unknown piece. This experiment can even be made more efficient by creating sticky ends using a six-pair base cutter (*SacI* in our case). Both approaches were used and primers were used that did not bind on the splicing variant. Bands of the correct size were obtained in many experiments but none of them was the expected FUT8.

4.2.4.7 FUT8-enrichment of the cDNA library for next-generation sequencing

One approach to get the whole sequence was to enrich the cDNA for Fut8 and sequence it with a next-generation sequencing approach. For enrichment either a biotinylated primer or a biotinylated probe are hybridized with the cDNA library and then separated by streptavidin coupled with magnetic beads. The enriched cDNA needs then to be further amplified by a PCR step to create a full double strand. To test for enrichment a PCR was done using specific primers for FUT8. No bands were visible after PCR on the enriched cDNA (Figure 38). Some optimization steps on washing and hybridization were also tried without success.

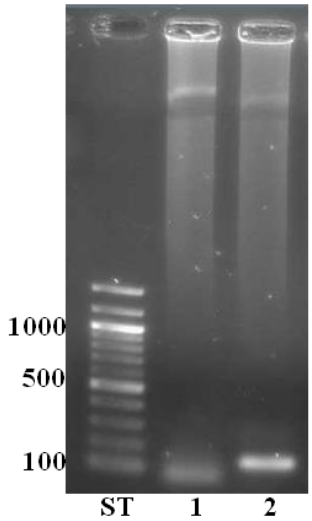


Figure 38: PCR using specific primer to prove enrichment of the library. Either when using the 200 bp probes or the primers no bands were visible after PCR. ST: 100 bp ladder 1: library enriched with biotinylated primers, 2: library enriched with biotinylated probes.

4.3 Establishing an enzyme activity assay for FUT8

The preferred acceptor substrate for so far known FUT8s is a GnGn oligosaccharide (Longmore & Schachter, 1982), which was prepared starting from fibrin through several steps. The goal was to establish a non radioactive enzyme activity assay using HPLC or MALDI and to use the acceptor for testing activity in the self-prepared microsomes. The main problem is that the oligosaccharide acceptor can not be labelled with 2-AP and purified through preparative HPLC, as the labelling destroys the structure of the reducing end N-acetylglucosamine (GlcNAc) which is recognized by the enzyme. The labelling is only possible after the enzyme activity assay.

Radioactive assays were also performed but they are very laborious as they need to be done in triplicate. Therefore, it is not an adequate method for testing activity for example during purification where a lot of fractions are tested for activity. In addition, when using radioactivity, it is not possible to discern between α-1,6 and α-1,3 fucosylation.

In this work most of the enzyme activity assays were measured by MALDI-TOF-MS as just one measurement was needed.

4.3.1 Preparation of the acceptor substrate for FUT8

Bovine fibrin was first digested with pronase, a mixture of several nonspecific endo- and exoproteases that digest proteins down to very short peptides. To separate the oligosaccharide containing peptides from the other peptides and the pronase and in order to desalt it, the mixture was applied to a gel filtration column. Fractions were measured at 280 nm for protein content and analyzed with orcin for sugar content (Figure 39).

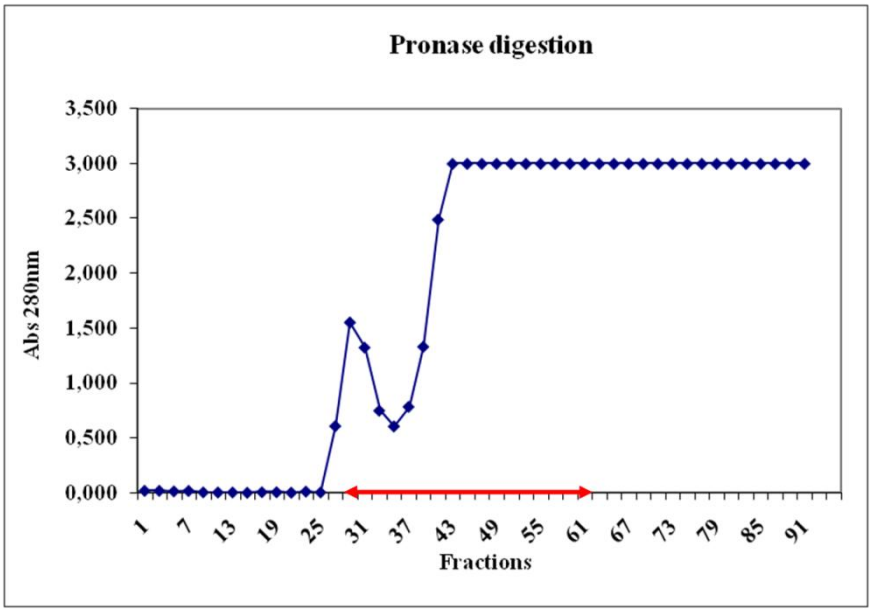


Figure 39: 280 nm measurement and orcinol spotting of the gel filtration fractions after pronase digestion. The first peek seen (fraction 26-37) is the elution of the pronase. The red arrowed line shows the pooled fractions (29-63).

In this case fractions 29 to 63 were pooled and lyophilised. The removal of sialic acids was done chemically by dissolving the lyophilised solution in 0,05 M H_2SO_4. After incubation in heat, the solution was applied to a gel filtration column and fractions analyzed by spectrophotometry and orcinol (Figure 40).

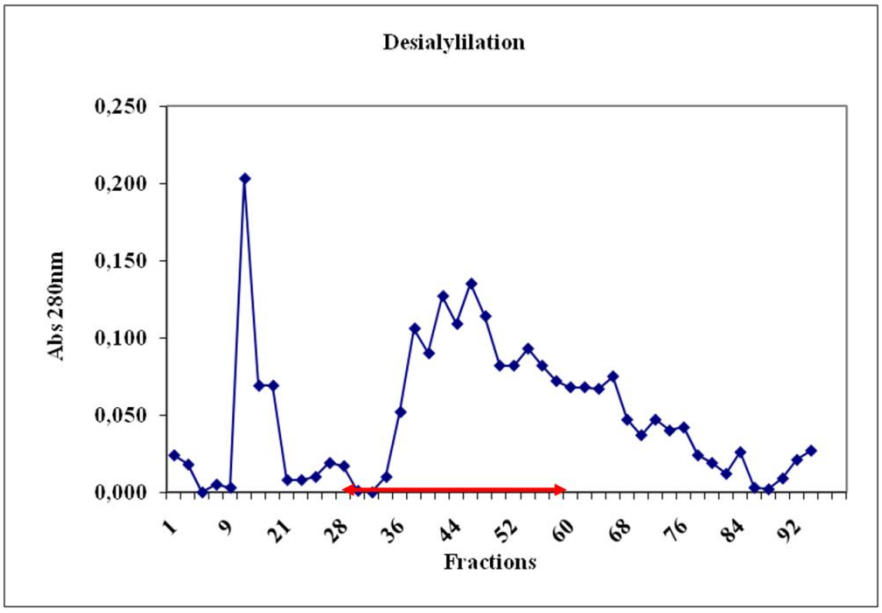


Figure 40: 280 nm measurement and orcinol spotting of the gel filtration fractions after desyalilation. The red arrowed line shows the pooled fractions (29-60).

In this case fractions 29-60 were pooled, lyophilised and then treated with PNGase A which should completely release the oligosaccharides from the peptides. After incubation, the solution was applied to a gel filtration column and fractions analyzed by spectrophotometry and orcinol (Figure 41).

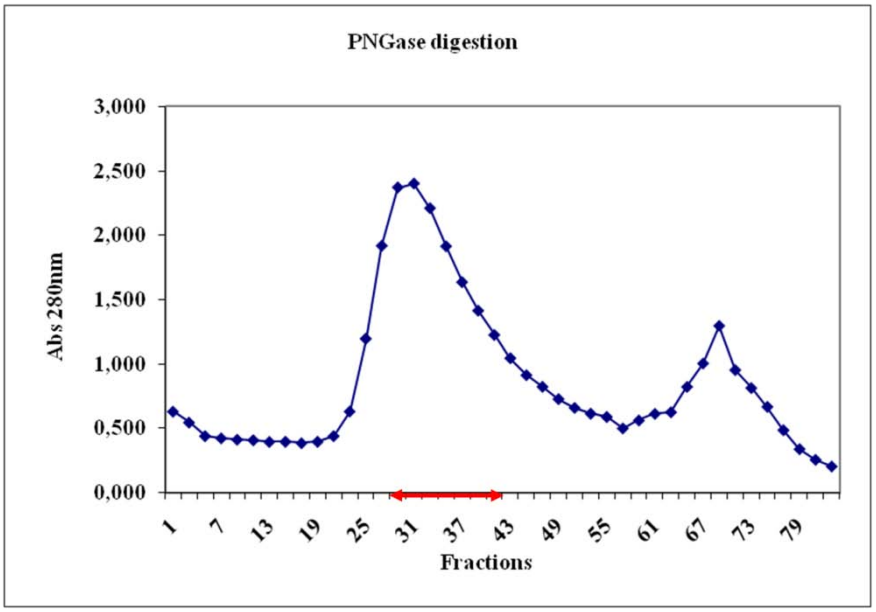


Figure 41: 280 nm measurement and orcinol spotting of the gel filtration fractions after PNGase A digestion. The red arrowed line shows the pooled fractions (29-43).

In this case fractions 29-43 were pooled and the volume reduced to approximately 2 ml. The pH was adjusted to 2 with concentrated acetic acid and the solution was applied to a cation exchange column. The free oligosaccharides should flow through the column whereas the ones still bound to peptides should be retained and elute by applying 0,4 M NH_4Ac pH 6.0. The fractions were analyzed by orcinol.

Fractions 3-11 were pooled and lyophilised. Fractions 19-31 were again treated with PNGase A. The free oligosaccharides were treated with galactosidase to remove the terminal galactoses and the sample was applied to a gel filtration column. Sugar content of the fractions was analyzed by orcinol.

Fractions 28 to 42 were lyophilized and the acceptor was then dissolved in H_2O and quantified. The concentration was 11,3 mM, which means that in 1 ml there were **11300 nmol**.

Especially the last enzymatic step with the galactosidase is not very efficient and sometimes one or both galactoses are not cleaved, therefore a quality control is needed prior to use. The quality control is performed by labelling an aliquot of the prepared acceptor with 2-AP which is then analyzed by Palpak Type S HPLC. This column separates the sugars by size. As a standard an isomaltose standard 3-11 glucose units and a purified 2-AP labelled GnGn oligosaccharide are run (Figure 42A, B and C).

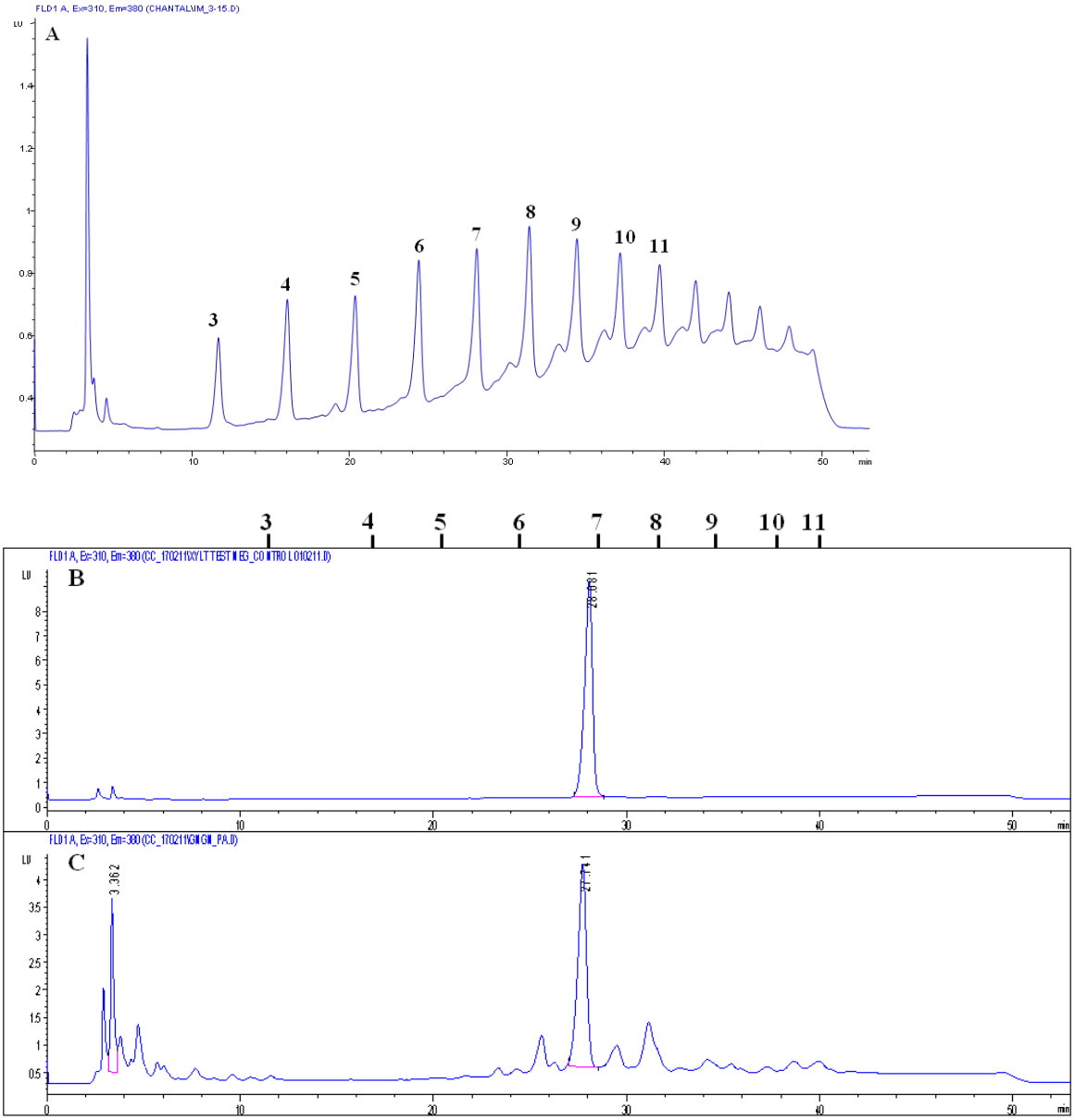


Figure 42A, B and C: Chromatograms of the isomaltose standard with peaks 3 to 11 indicating the glucose units (A), purified GnGn oligosaccharide control (B) and new acceptor sample (C). The numbers above B and C indicate the glucose units.

Some oligosaccharides carry still one galactose, but the quality is sufficient for the enzyme activity assay of FUT8.

4.3.2 Enzyme activity assay with homogenised rabbit brain

In order to establish a non-radioactive assay for FUT8, the assay conditions were tested with a homogenised rabbit brain preparation where the activity has already been confirmed (Struppe & Staudacher, 2000). As negative control an assay without GDP-fucose was made. The preparations were then labelled with 2-AP and analyzed on the Palpak Type S HPLC. In parallel the GnGn oligosaccharide acceptor used for the assay was run as an additional control (Figure 43A, B and C).

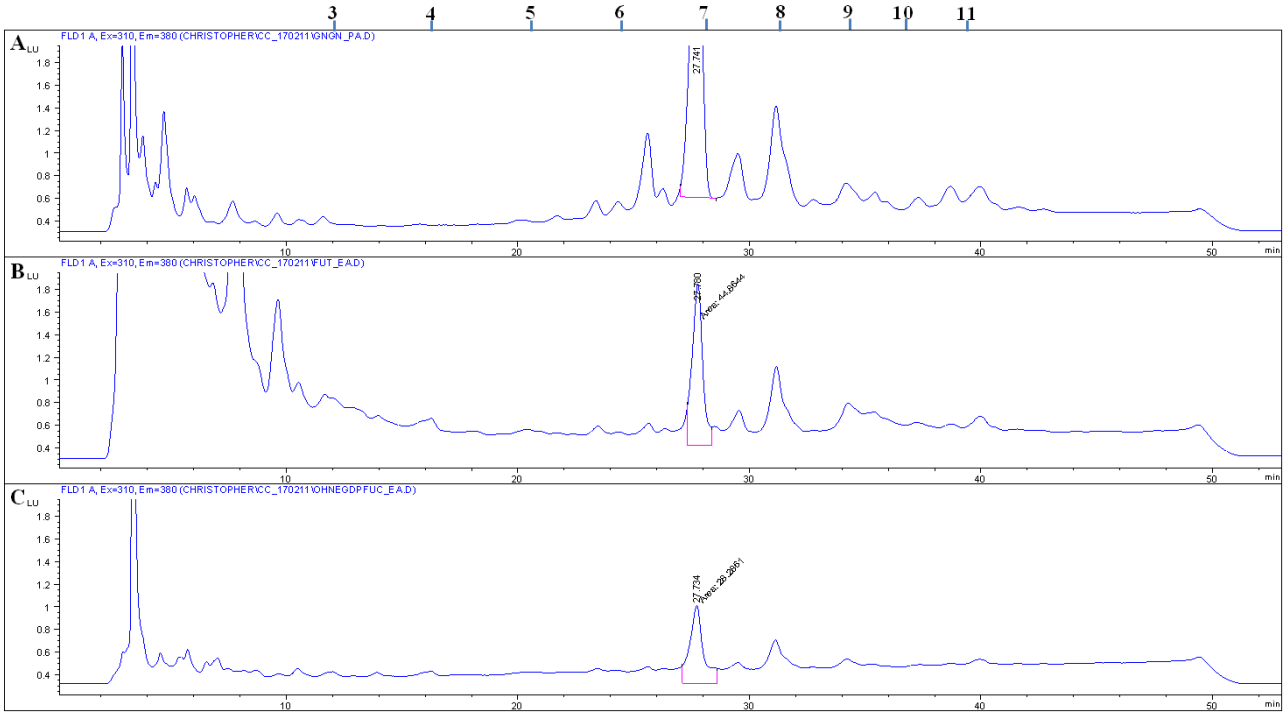


Figure 43A, B and C: Chromatogram of the enzyme activity assay. In (A) GnGn oligosaccharide, in (B) the enzyme activity assay with GnGn oligosaccharide and in (C) the negative control without GDP-fucose. The numbers indicate the glucose units.

The activity was not visible on HPLC as the peak with one additional glucose unit (8) did not increase.

4.3.3 Microsome preparation of *A. lusitanicus*

As FUT8 is located in the Golgi, one possibility is to prepare microsomes to get rid of enzyme activity disturbing factors and enrich the enzyme. The microsome preparation of *A. lusitanicus* was then tested for activity and analyzed with Palpak Type S HPLC (Figure 44A, B and C).

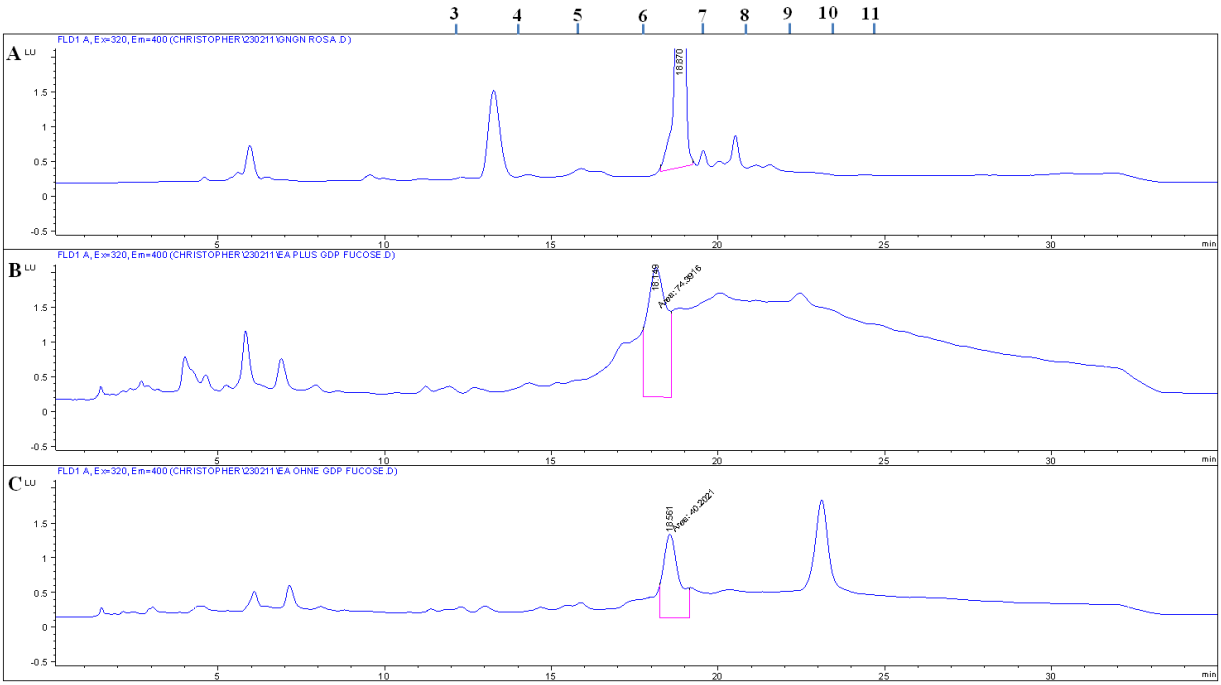


Figure 44: Chromatogram of the enzyme activity assay. In (A) GnGn oligosaccharide, in (B) the enzyme activity assay with GnGn oligosaccharide and in (C) the negative control without GDP-fucose. The numbers indicate the glucose units.

No enzyme activity was visible.

4.3.4 Microsome preparation of Bge cells

Similarly to *A. lusitanicus* microsomes were purified from Bge cells as they are an easily available source which is much less complex than the whole organism. This time the activity was analyzed by a radioactive enzyme activity assay. Activity was measured for the whole cells homogenate, cell pellet, cell medium and microsomes. For the negative control the acceptor was omitted. Even if there is obviously some enzyme activity in the microsomes, the standard deviation is huge and therefore the result not very trustable. This was a common problem in several experiments of the same kind (Figure 45A and B).

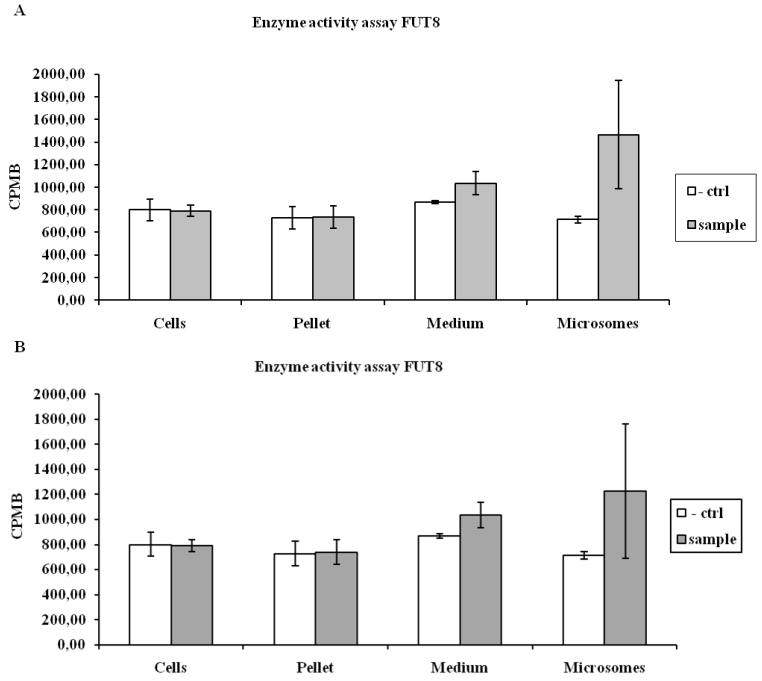


Figure 45A and B: Enzyme activity of Bge cells, cell pellet, medium and microsomes of two preparations (A and B). Left: without acceptor, right: with GnGn oligosaccharide as acceptor.

4.4 Cloning and expression of a soluble incomplete form of *A. lusitanicus* FUT8 (amino acid residues 29 to 507)

We were not able to find the 3'end of *A. lusitanicus* FUT8 corresponding to the complete SH3 domain and to the C-terminus. Therefore, as only approximately 130 bp (~ 43 amino acids) were missing, our strategy was to express the soluble form of *A. lusitanicus* FUT8 without the missing domaint and check for its activity. Two expression systems were chosen that are usually used for glycosyltransferases, *P. pastoris* and insect cells. For *P. pastoris* cells the vector of choice was a modified version of the pPICZα vector, pPICZα FlagHis1, as it contains the sequence of the α-mating factor from *Saccharomyces cerevisiae* for secretion, a 6xHis-Tag for purification and a Flag-Tag for detection, both on the N-terminus. The insect cell vector was pVT-Bac which also provides a 6xHis-Tag on the N-terminus and features the melittin secretion signal from honeybee. Expression was analyzed by SDS PAGE and WB. Enzyme activity was measured by MALDI-TOF.

4.4.1 Cloning FUT8 in pVT-Bac and pPICZα FlagHis1

A. lusitanicus Fut8 was amplified starting from the directional cDNA library (Figure 46), digested with *KpnI* and *PstI* and cloned in both vectors.

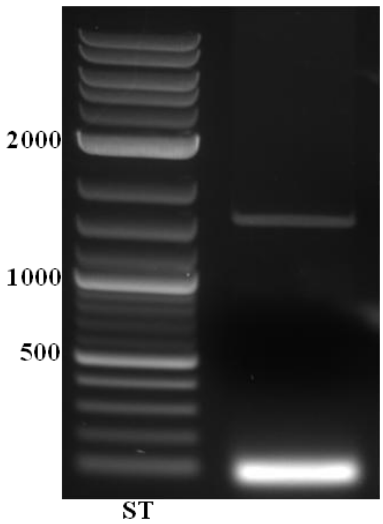


Figure 46: FUT8 from *A. lusitanicus* directly amplified from the directional library. ST: 2log ladder.

After transformation in *E. coli* cells, PCR screening was performed and the positive clones sent for sequencing. One clone in pPICZα FlagHis1 and two clones in pVT-Bac had the correct sequence and were used for further experiments.

4.4.2 Transformation and expression in *P. pastoris* cells

The plasmid containing the correct clone was first digested with *PmeI* before transformation in *P. pastoris* cells which was performed with electroporation. After 48 h of incubation eight *P. pastoris* clones were picked and PCR screening was performed in order to check the correct integration of the sequence into the genome (Figure 47). A control PCR was performed in parallel using primers for a housekeeping gene.

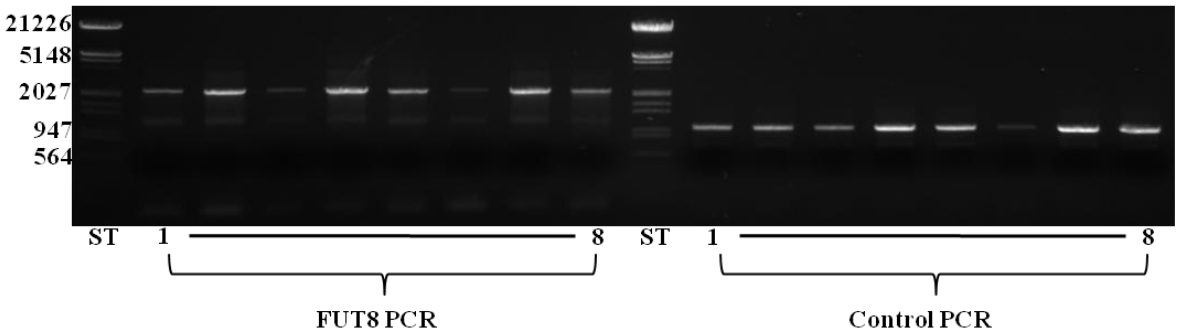


Figure 47: PCR screening and control PCR of eight *P. pastoris* clones. ST: λ DNA HindIII/EcoRI ladder, 1-8: screened clones. All clones are positive.

In all clones the sequence integrated correctly in the genome of *P. pastoris* cells. From six of the eight clones a 5 d expression was started. The efficiency was analyzed by Coomassie and WB of pellet and supernatant using Penta-His monoclonal antibody (Figure 48A and B).

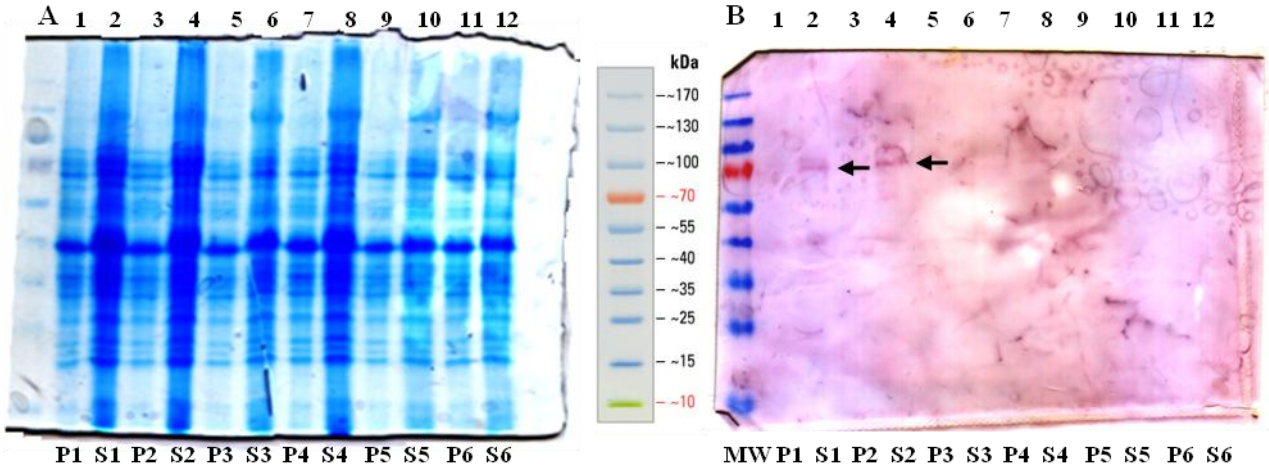


Figure 48A and B: Coomassie (A) and WB using Penta-His antibody (B) of the six *P. pastoris* clones showing in the WB faint positives band (indicated by the arrows) on 1S and 2S. MW: molecular weight ladder, 1-6: clones; S: supernatant; P: pellet.

Two of the clones (1 and 2) seemed to express the FUT8 and secrete it into the supernatant (lane 2 and 4 of the WB), but the band was a little bit higher as expected: the enzyme is approximately 60 kDa and the band appears at 70 kDa. The blot above was done using Penta-His monoclonal antibody as the Anti-Flag antibody always showed an unspecific band exactly where the protein is expected.

The supernatant of clone 2 was tested for activity by MALDI-TOF but no activity was seen.

4.4.3 Cotransfection and expression of the incomplete FUT8 expressed in insect cells

One of the two plasmid preparations of the correct clone was used for cotransfection in *Sf*9 insect cells. After two rounds of scale-up, the infectious supernatant was used for expression experiments in *Sf*9, *Trichoplusia ni* High Five cells or *Ascalapha odorata Ao*38 cells.

In a first experiment expression in *Sf*9 cells was compared to expression in *Trichoplusia ni* High Five cells. Coomassie and WB of supernatant and pellet of cells expressing FUT8 and a His-Tagged control protein were made (Figure 49A and B).

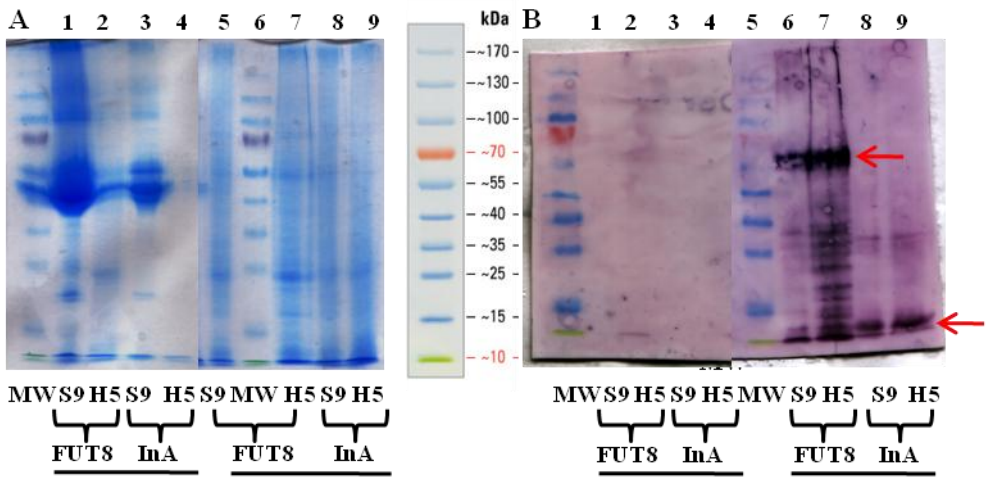


Figure 49A and B: Coomassie (A) and WB using Penta-His antibody (B) of *Trichoplusia ni* High Five (H5) and *Sf*9 (S9) cells expressing FUT8 and a His-tagged control protein (InA). MW: molecular weight ladder.

As a melittin sequence is on the plasmid it was expected that the protein is secreted in the medium but the WB showed that in both cell lines it was retained in the cell (lane 6 and 7 on the WB). This was also true for the control protein which was already known to be in the pellet (8 and 9 of the WB)rather than in the supernatant. There was no real difference between the cell lines, but *Trichoplusia ni* High Five cells were preferred for the next experiments as they can be cultivated without FCS.

In the next experiment expression in *Trichoplusia ni* High Five cells was compared to expression in *Ascalapha odorata Ao*38 cells hoping that they may secrete FUT8. Here, just the WB is showing again that the enzyme stays in the cell pellet (Figure 50).

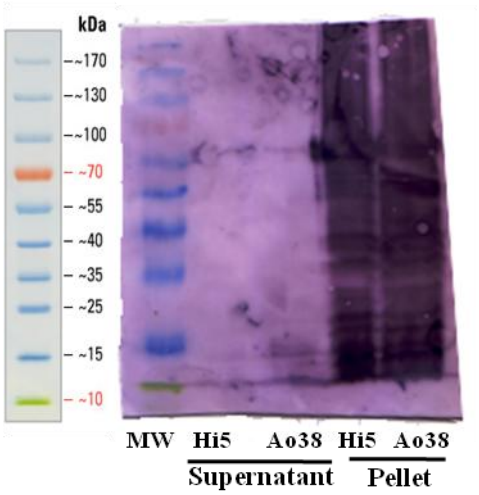


Figure 50: WB using Penta-His antibody of supernatant and pellet of *Ascalapha odorata Ao*38 (Ao38) and *Trichoplusia ni* High Five (Hi5) cells expressing FUT8. In both cell lines the enzyme stays in the cell. MW: molecular weight ladder.

As a last experiment a time curve in *Trichoplusia ni* High Five cells was made in order to check at which day after infection the expression of FUT8 was optimal. Again supernatant and pellet of day 2, 3 and 4 after infection were analyzed by Coomassie and WB (Figure 51A and B).

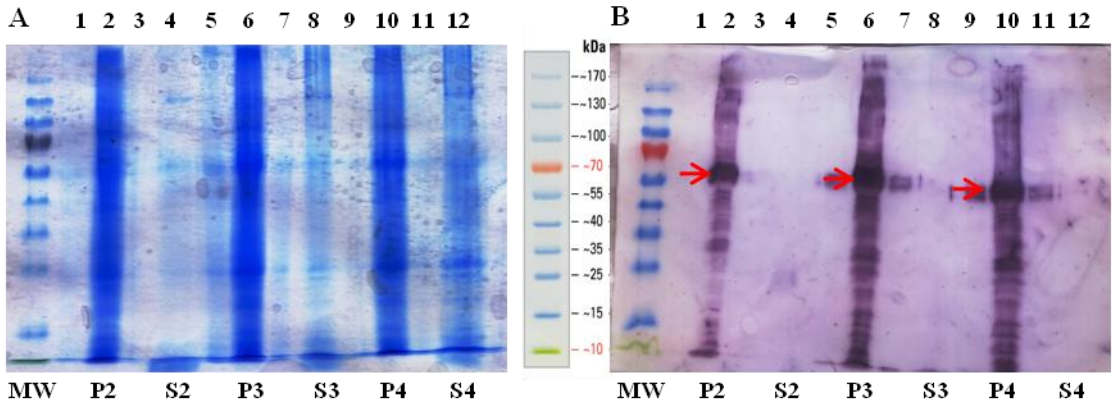


Figure 51A and B: Coomassie (A) and WB using Penta-His antibody (B) of supernatant (S) and pellet (P) of *Trichoplusia ni* High Five cells expressing FUT8 after 2, 3 and 4 days after infection. MW: molecular weight ladder.

There was no difference in the protein quantity between the days and as expected the enzyme was in the pellet (lane 2, 6 and 10 on the WB).

The pellet was lysed by sonication and the supernatant was tested for activity by MALDI-TOF but no activity was seen.

4.5 Cloning and expression of four soluble FUT8 fusion proteins

The missing ~43 aminoacids at the C-terminal of *A. lusitanicus* FUT8 may be responsible for the lack of activity. Some of these amino acids belong to the highly conserved SH3 domain and even if its role is still not well understood, it might be important for functionality. Therefore, we constructed four different fusion proteins using the known soluble part of the *A. lusitanicus* FUT8 and replacing the whole SH3 domain with the SH3 domain of *Crassostrea gigas*, *Lottia gigantea*, *Caenorhabditis elegans* and *Drosophila melanogaster* respectively. The first two organism are the evolutionary nearest one to the slug but the activity of their enzymes has until now not been proven. The other two are the evolutionary nearest organism were the activity of FUT8 was confirmed (Paschinger et al, 2005b).

4.5.1 Design of the constructs

The fucosyltransferase sequences except the one of *L. gigantea* were obtained from GenBank. *L. gigantea* sequence was obtained by blasting the known *A. lusitanicus* sequence on the *L. gigantea* project homepage http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html.The SH3 domains were identified using the conserved domains tool from NCBI http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi.

As example one of the constructs is shown in (Figure 52).

```
GTAGTActgcagGATGCAGCTAATCATGCAGAACGTAGCTTGTCAAGAGCCATGGAAGAACTGGATAA
ATTGTACCAGCAAAATAAGAAACTTGAGCAACTTATAATGCAACTAAAAAAAAATGAAACTTTTCCTC
ATGGCCTTCGGGACAATGTTACTGTAGAGAAACTAGAATTCCGACTGTTGAGAGCTAGCAATGAACTG
TCTCAGATAGCTGATTCTACACAGACTAAGTTAACATATGAGGGTGAACAAGCCAGAAGAAAGGCAGA
GAACACAGTGAAAGAATTGTGGTACTTTTTAAACTCTCAGCTCAAGAAACTTGACCACATAGACAACA
ATGATAATCATTTGACAGAAAGAATATCTAAGCTGAAGAAAGATCTGGAAGGATATCGACGGACTACT
TTAGAGGATTTTGAGAAGCTACGCAAGACAAACCATGCAGACGATTATAGGCTACAGAAGTCACAAGA
GCTTGGAGATCTGGTTCAGAGAAGACTGGAATATATACAAAACCCAGTCAACTGTAAGACTGCAAAAA
AAATTGTTTGCAACTTGCACAAAGGATGTGGTTTTGGCTGTCAGCTCCATCACATCACCTACTGTCTT
ATAGCAGCATATGCCATGGAAAGAACTCTGATACTTGATTCCAAAGGCTGGAGGTACTCACCCACAGG
CTGGGAGTCTGTGTTCGAACCTCTTAGCAAGACTTGTTCACAAGTAAACAATGAAGGTTCCAGAACCC
ACTGGAGATCATCTGTGGAAGAAATGAGCTCGTATGACATCATTGACCTTCCTATCGTTGACAGCCTT
CATCCTCGTCCGGGTTTTATGCCACTCTCAGTGCCCGCAGATCTTGCTCATGACATCTCAATATTTCA
TGGTGACCCGTCGGTGTGGTGGATAGGTCACATTGTTCGTTACTTGTTCAGATTAAGACCAATGGTAC
TTCAGGATGTGGTCAATGCTGGCAAGAAGATGGGATTCCAAAACACCATTGTTGGGGTACATGTGCGT
AGAACAGATAAAATTGATTTGGAAGCAGCCTTCCATCCCCTTTTGGAATACATGCTACATGTCGGGGA
GTACTTTGACCAGTTGGAGAGAACACAGTCTAATATCACTAGGAGGGTATACCTGGCCTCAGACGACC
CAAATGTACTTACAGAGGCCCAGAAACAATATCCAGCGTATAT
```

```
GTTTATCAGTGACAGGTCTATATCGCAGTCTGCAAGCTTAGGTACTAGATATACCGACAACTCATTAC
GAGGGATTGTTATTGATATTTACTATCTTGCTCGTTGCGACTTCCTGGTCTGCACGTTTTCATCGCAG
GTTTGCAGAGTTGCCTATGAGATTATGCAGACATTGCATGGAGATGCATCGAAAAACTTCCGCTCACT
TGATGACATCTTCTACTATGGAGGACAGAATGCACATGACCTGGAGGCAGTGGAAAAACATGTCAAAC
AGAACGAAAAAGAAATAGACTTGGAGCCTGGCGACTTGGTGGGCATTGCTGGAAATCACTGGGATGGT
TATTCTAAAGGCATGAACCACCGAACAGGGAAAACGGGGCTATTCCCCTCATATAAAACTAGAGAAAA
ATACACCATAGTGGATTTACCAACATATCCAGAAGTTTCAGAAGGCGGATGAggtaccATGCGTC
```

Figure 52: FUT8 fusion protein. The known soluble part of *A. lusitanicus* and the SH3 domain and C-terminus of *C. gigas* is in grey. In yellow is the *PstI* restriction site and in green the *KpnI* restriction site for cloning into the pVT-Bac vector. In red is the *HindIII* restriction site for attaching the gBlock (in the box) to the known sequence. Underlined in the red *HindIII* restriction site is the silent mutation necessary for creating the restriction site.

4.5.2 Cloning of the constructs in pVT-Bac

In order to ligate the SH3 gBlocks to the *A. lusitanicus* sequence a silent mutation was created using a PCR on a plasmid preparation with the pVT-Bac-FUT8 construct and a primer containing the mutation. Then the PCR product was digetsed with *PstI* and *HindIII*: In parallel, the gBlocks were digested with *HindIII* and *KpnI*. PCR products and gBlocks were ligated together and then ligated to the already linearized pVT-Bac. After transformation in *E. coli* cells, a PCR screening was performed (Figure 53).

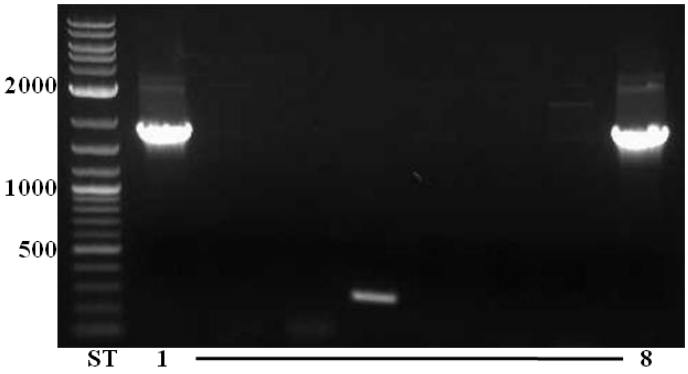


Figure 53: PCR screening of the fusion protein containing the known soluble part of *A. lusitanicus* and SH3 domain and C-terminus of *C. gigas*. ST: 2log ladder, 1-8: screened clones. Clones 1 and 8 are positive.

Only the PCR screening of the *C. gigas* fusion protein gave positive clones. The plasmid preparations of the two clones were subjected to sequence analysis and just clone 8 was suitable for cotransfection in *Sf*9 cells.

After two rounds of scale-up, the infectious supernatant was used for expression in *Trichoplusia ni* High Five cells. The expression was analyzed as usual on a Coomassie stained SDS PAGE and WB using Penta-His antibody. First, just supernatant and pellet were applied to the gel and analyzed on the WB (Figure 54A and B).

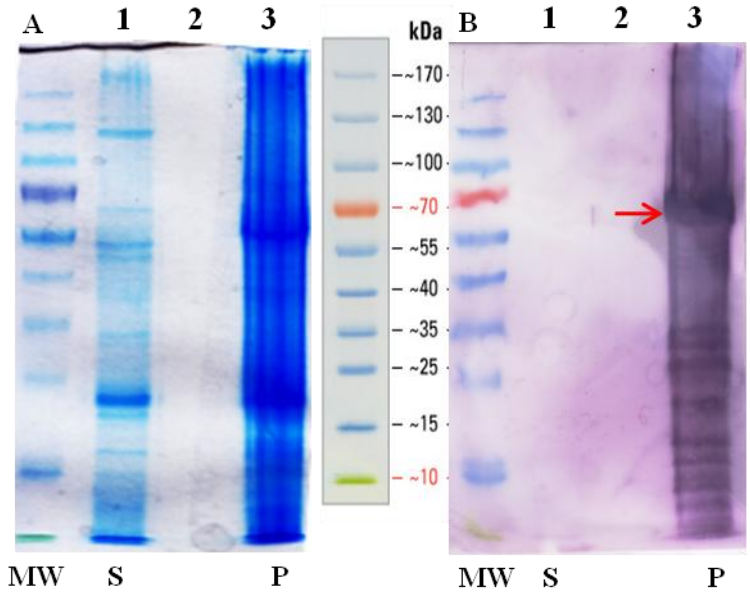


Figure 54A and B: Coomassie (A) and WB using Penta-His antibody (B) of supernatant (S) and pellet (P) of *Trichoplusia ni* High Five cells expressing the *C. gigas* fusion protein. MW: molecular weight ladder. As expected, the protein is again located intracellular.

As already seen for the incomplete FUT8, the enzyme is located in the cell (lane 3 on the WB) and not secreted even when the melittin secretion signal is present on the vector. The pellet was lysed by sonication and soluble and insoluble part of it were applied to the gel and analyzed by WB. On the following figures, also the supernatant of *Sf*9 cells expressing ppGalNAcT from Bge cells is shown (lane 1 on the Coomassie and WB), as it was applied for an expression control (Figure 55 A and B).

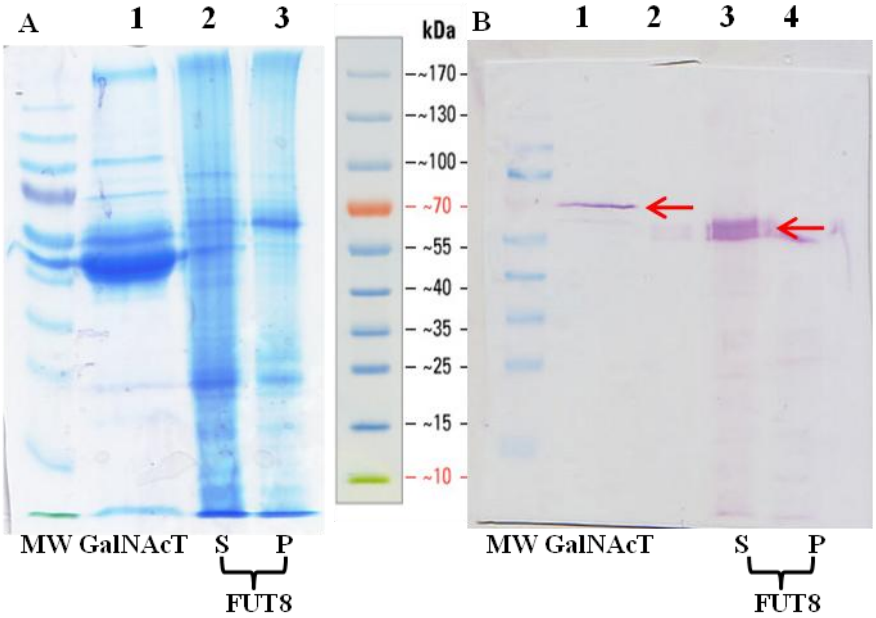


Figure 55A and B: Coomassie (A) and WB using Penta-His antibody (B) of soluble (S) and insoluble (P) part after lysis of *Trichoplusia ni* High Five cells expressing the *C. gigas* fusion protein. MW: molecular weight ladder, ppGalNAcT: polypeptide N-acetylgalactosamine transferase (GalNAcT) from Bge cells; FUT8 is in the soluble part.

The enzyme was located intracellular but it is still soluble (lane 3 on the WB). However no enzyme activity was seen on MALDI-TOF.

4.6 Cloning and expression of *Lottia gigantea* and *Crassostrea gigas* FUT8s

As neither the incomplete FUT8 nor the complemented were active, we tried to express the complete FUT8s from *L. gigantea* and *C. gigas* as the sequences were available. The whole genome of *C. gigas* has been sequenced, which is the first mollusc genomic sequence available (Zhang et al, 2012). Both genes were synthetically synthesized.

The sequences were digested with *PstI* and *KpnI* and cloned into pVT-Bac vector for cotransfection in *Sf9* cells. PCR screening was performed (Figure 56).

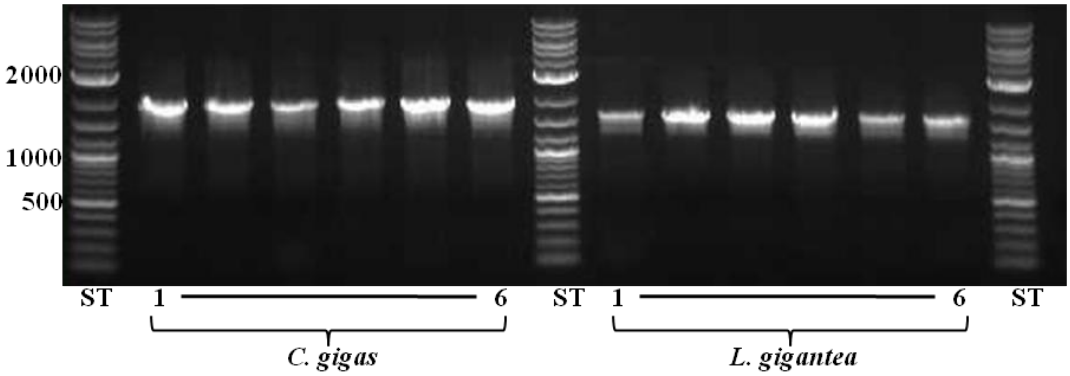


Figure 56: PCR screening of FUT8 from *C. gigas* and *L. gigantea* cloned in pVT-Bac. ST: 2log ladder, 1-6: clones tested. All clones are positive.

The expression was analyzed as on a Coomassie stained SDS PAGE and WB using Penta-His antibody. Supernatant and the soluble and insoluble part after cell lysis were applied to the gel and analyzed with WB (Figure 57A and B).

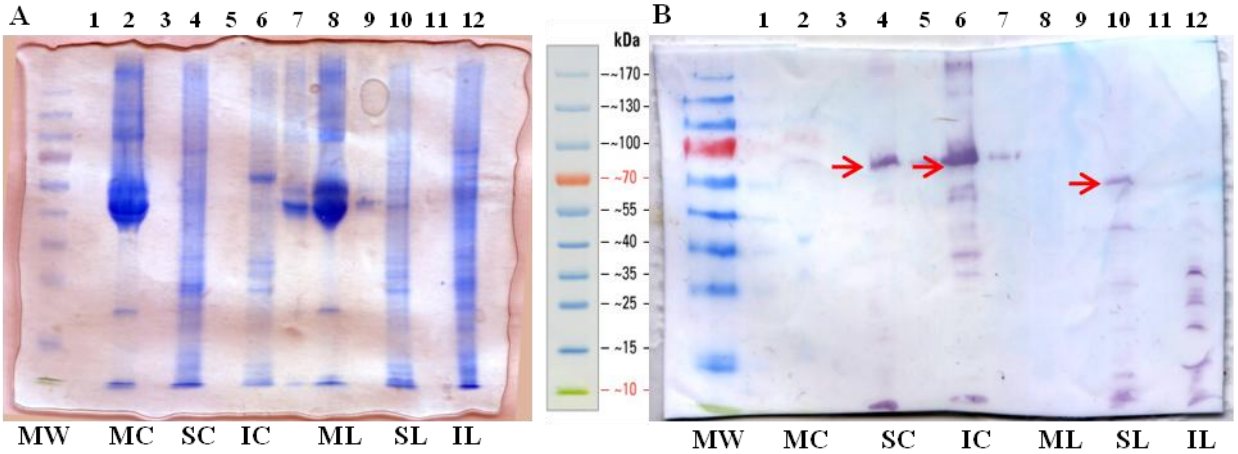


Figure 57A and B: Coomassie (A) and WB using Penta-His antibody (B) of the expression of *Crassostrea gigas* (C) and *Lottia gigantea* (L) FUT8 in *Trichoplusia ni* High Five. Medium (M) and the soluble (S) and insoluble (I) part after cell lysis were analyzed. The *C. gigas* FUT8 is found in the soluble and insoluble part whereas in *L. gigantea* it seems to be just in the soluble part.

Both enzymes were not secreted as already seen for the different *A. lusitanicus* constructs. *C. gigas* FUT8 was expressed in a soluble and insoluble form (lane 4 and 6 on the WB) whereas in *L.gigantea* it seems to be just in the soluble part (lane 10 on the WB).

No enzyme activity was seen on MALDI-TOF.

4.7 ß- galactosidase purification from *A. lusitanicus*

In a side project the biochemical purification of a ß-galactosidase from *A. lusitanicus* was performed. The purification protocol was pretty well established but a real purification of the enzyme was never achieved and the protein sequence is still unknown. This time we wanted to have enough material to get the protein sequence and design degenerate primer to amplify the whole sequence from the cDNA library.

The purification started by removing the gut from the slugs. After homogenization of the samples, two ammonium sulphate precipitation (40 and 80%) were performed and the resulting pellet is dissolved in starting buffer for hydrophobic interaction chromatography.

4.7.1 Hydrophobic interaction chromatography

The hydrophobic interaction chromatography step helped to separate hydrophobic (e.g. transmembrane proteins) from hydrophilic proteins (e.g. exoglycosidases). Fractions were collected and measured at 280 nm for protein content and enzyme activity (Figure 58).

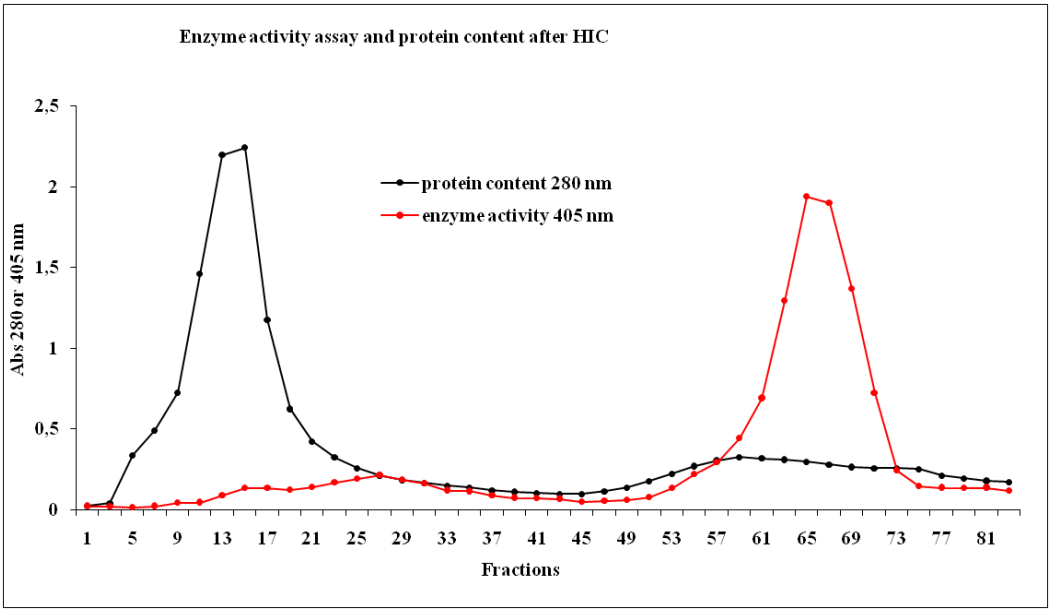


Figure 58: Protein content (black) and enzyme activity (red) after HIC.

The protein content peak confirmed that the hydrophobic proteins were efficiently separated from the ß-galactosidase, which is a hydrophilic protein.

In this case fractions 55 to 75 were pooled, the buffer changed and the sample applied to an anion exchange chromatography column.

4.7.2 Anion exchange chromatography

Through anion exchange chromatography proteins with different charge were separated. The sample was applied onto a DE52 column and the column was washed with 2 CV. Then a linear gradient up to 1 M NaCl was applied. Fractions were collected and were measured at 280 nm for protein content and enzyme activity (Figure 59).

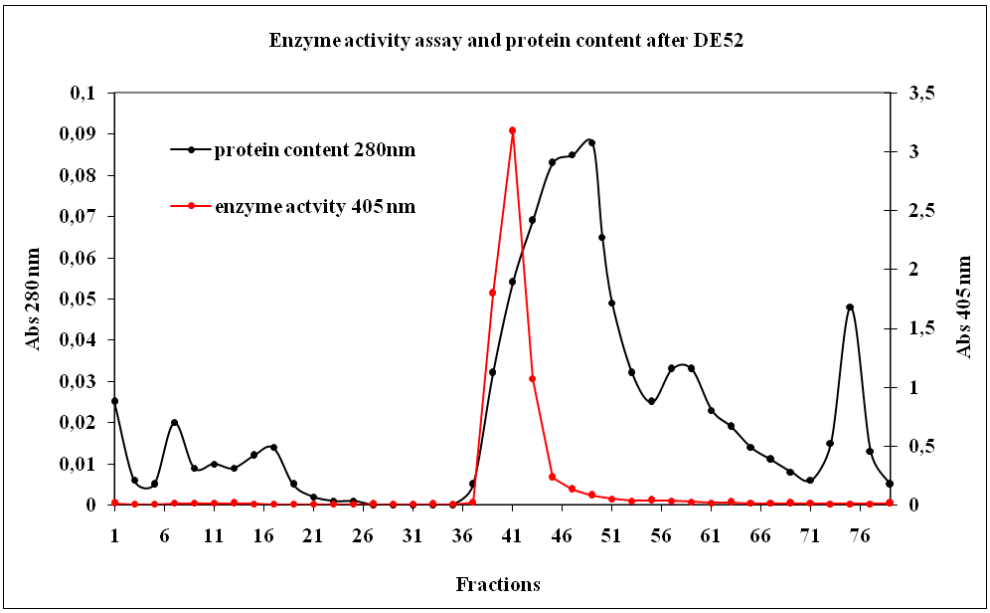


Figure 59: Protein content (black) and enzyme activity (red) after DE52.

Also here there was a good separation effect. In this case fractions 37 to 51 were pooled, the volume reduced and the sample applied to size exclusion chromatography.

4.7.3 Size exclusion chromatography

To separate big proteins from the smaller ones size exclusion chromatography was performed using a Sephacryl S-200 column. In this step we hoped to separate the ß-galactosidase from other exoglycosidases such as ß-N-acetylglucosaminidase. The sample was applied to the column and fractions were collected. Protein content was measured and starting with fraction 30 enzyme activity of different exoglycosidases (ß-galactosidase, α-glucosidase, ß-xylosidase, α-fucosidase, α-mannosidase and ß-N-acetylglucosaminidase) was tested (Figure 60 and Figure 61).

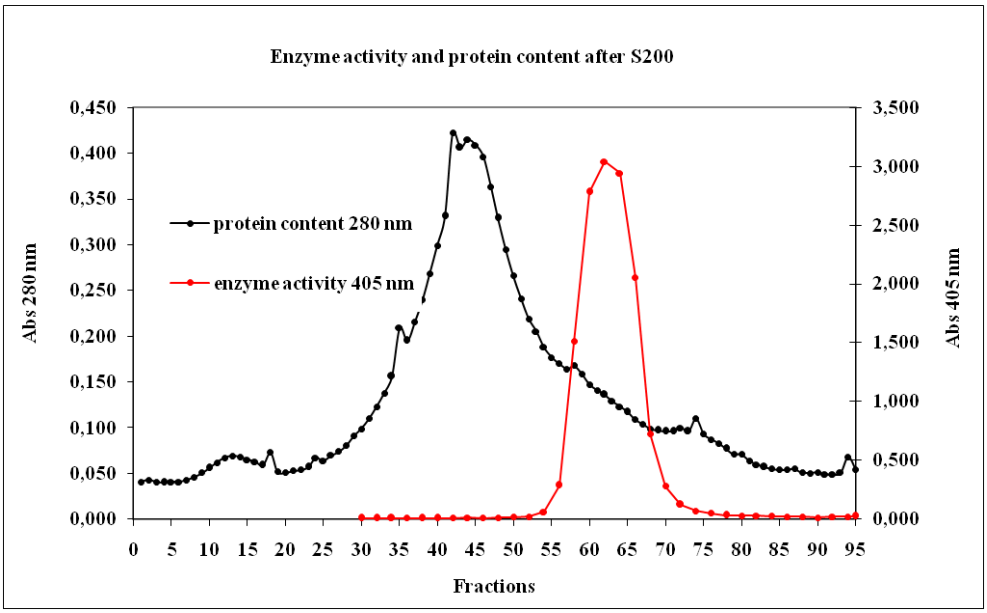


Figure 60: Protein content (black) and the enzyme activity (red) after S200.

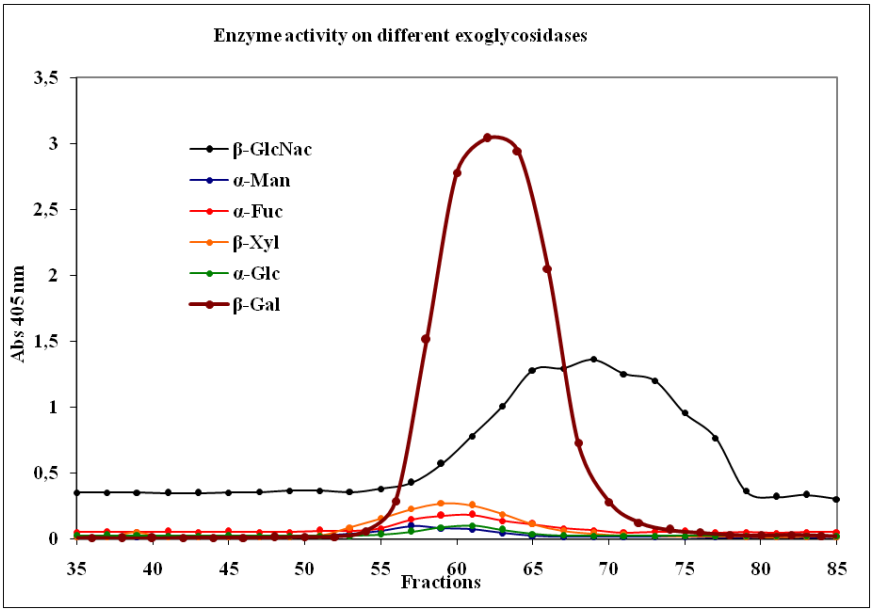


Figure 61: Activity of different exoglycosidases after S200.

The first figures show that it is possible to separate some bigger proteins from the ß-galactosidase. Most of the exoglycosidases co-elute with the ß-galactosidase but it is possible to get rid of at least some of the ß-N-acetylglucosaminidase (Figure 61).

In this case fractions 57 to 61 were pooled, buffer changed and applied to affinity chromatography.

4.7.4 Affinity chromatography

The last step was necessary to separate the co-eluted exoglycosidases from the ß-galactosidase. Therefore, two columns containing a substrate analogue for the ß-galactosidase were tested:

- ρ-aminophenyl ß-D-thiogalactopyranoside
- ρ-aminobenzyl 1-thio-ß-D-galactopyranosid

The columns were run under the same conditions and fractions were collected. From each fraction protein content and enzyme activity of different exoglycosidases (ß-galactosidase, α-glucosidase, ß-xylosidase, α-fucosidase, α-mannosidase and ß-N-acetilglucosaminidase) was measured (Figure 62A, B and C).

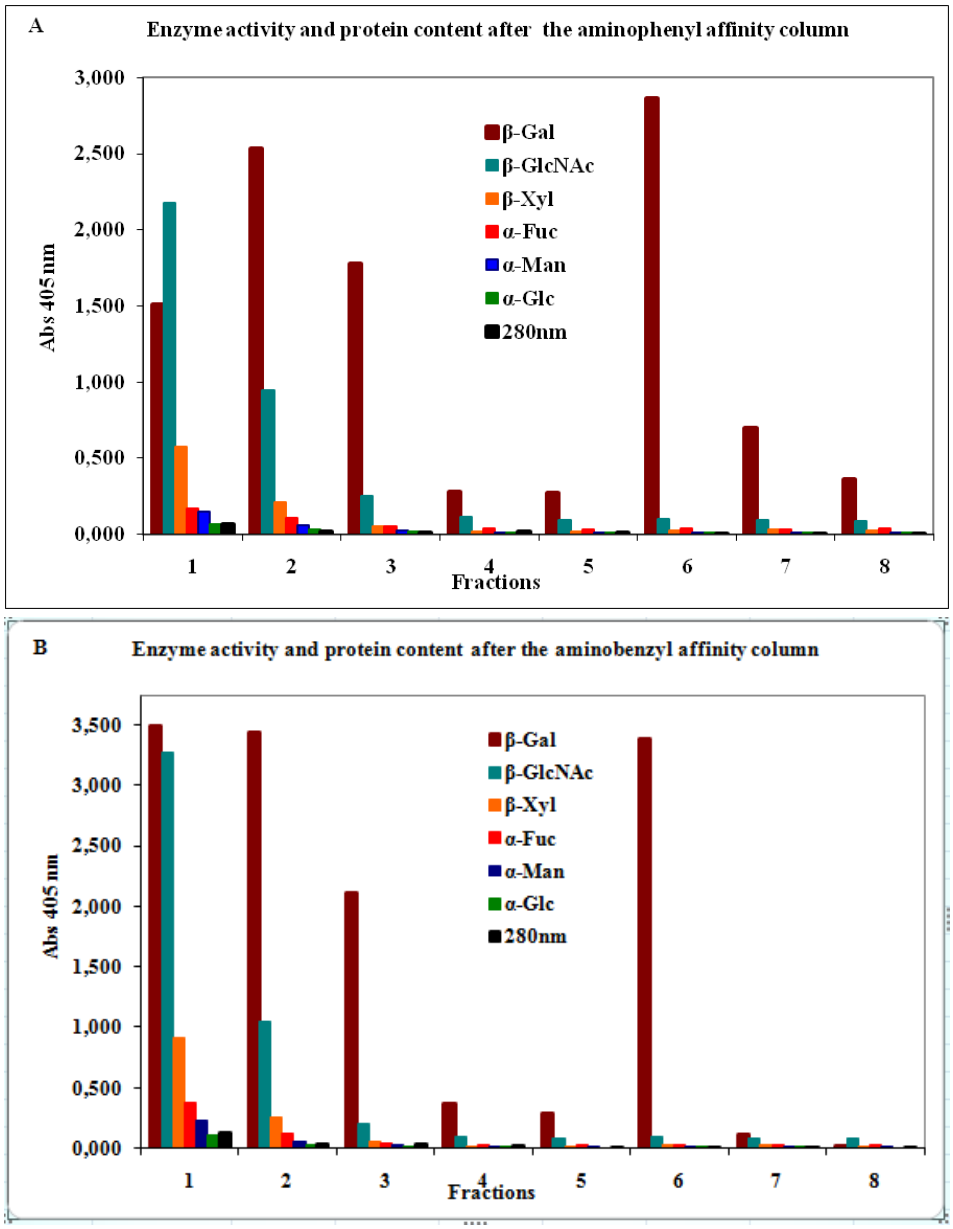


Figure 62A and B: Protein content and the enzyme activity of different exoglycosidases after the aminophenyl column (A) and the aminobenzyl column (B).

There was still a lot of ß-galactosidase which did not bind to both columns (see fraction 1-3). It is also visible that the purification with the aminobenzyl column was much better, as the ß-galactosidase was concentrated mainly in fraction 6 and there was nearly no activity of the other glycosidases in this fraction (Figure 62B). For further experiments just the aminobenzyl column was used. All fractions were analyzed with Coomassie and silver staining.

Fractions 1-3 were applied for two further rounds to the affinity columns and this time just ß-galactosidase activity was tested (Figure 63A and B).

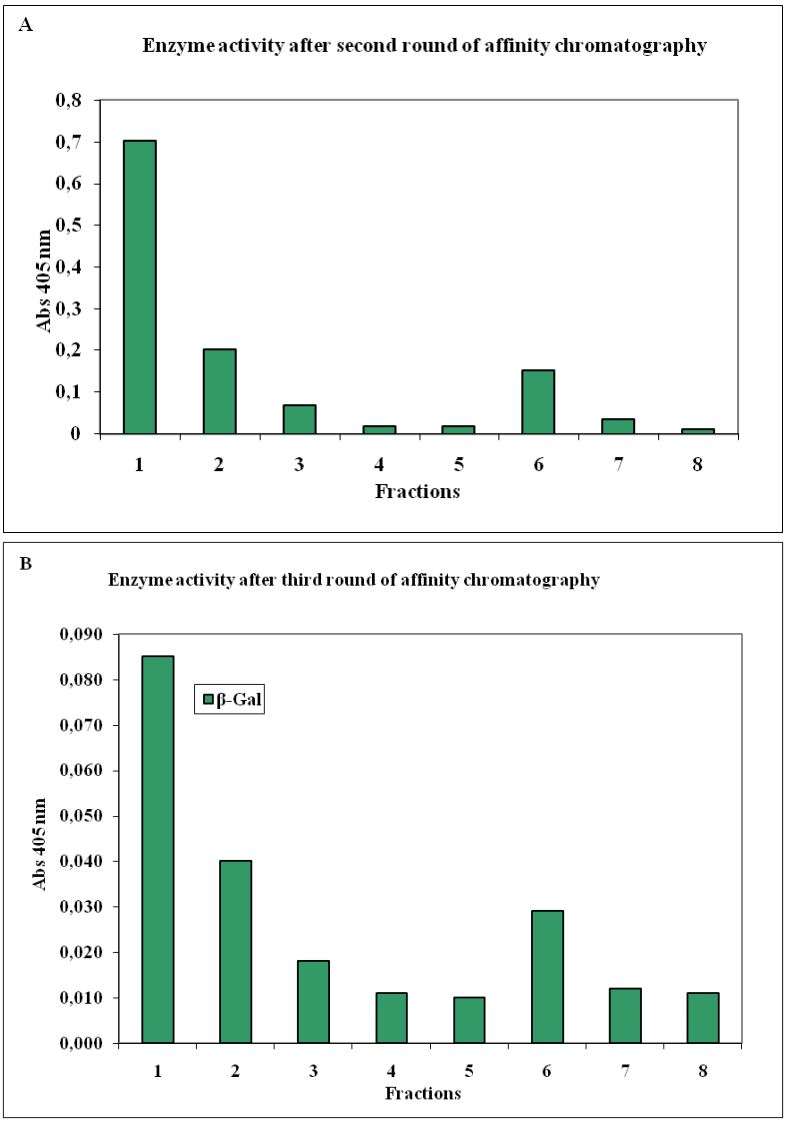


Figure 63A and B: ß-galactosidase enzyme activity of second (A) and third (B) round of affinity chromatography with the aminobenzyl column.

Even when some additional ß-galactosidase could be recovered, there was still some active enzyme which did not bind.

4.7.5 SDS-PAGE with Coomassie and silver staining of the last chromatography step

All fractions from the aminobenzyl column were methanol precipitated and applied to an SDS PAGE. First a Coomassie staining was performed and as only few bands were visible also a silver staining was performed (Figure 64A and B).

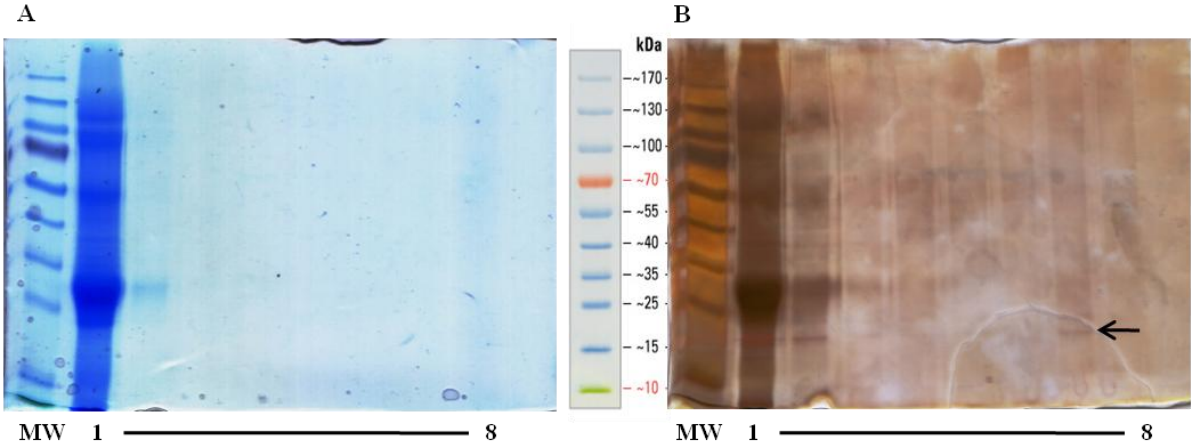


Figure 64A and B: Coomassie (A) and silver staining (B) of the aminobenzyl affinity chromatographic step. MW: molecular weight ladder, 1-8: chromatography fractions. Possible ß-galactosidase band is indicated with an arrow.

On the electrophoresis it is visible that with this step we could get rid of a lot of proteins. No bands are clearly visible, however in the silver staining there is a faint band in fraction 7 of approximately 20 kDa which could be the ß-galactosidase, nonetheless it is not enough for determination of the protein sequence.

Dr. Erika Staudacher managed to get enough protein to excise a band from the Coomassie stained gel and to get four peptide sequences. However these sequences did not match with any glycosidases in the databases.

Peptide number	Peptide sequence
1	LDYVVSSAEQHDIK
2	CPSCDTSVLYNWIEK
3	AAGKPCLLEEYGVTSNHCSVEGSWQK
4	TALSTTGVGADLFWQYGDDLSTGK

Table 3: Peptide sequences from the ß-galactosidase purification.

5 Discussion

N-glycans which are α-1,6 –fucosylated are found in nearly each organism of the animal kingdom (Staudacher et al, 1999) and also gastropods are no exception (Gutternigg et al, 2004; Gutternigg et al, 2007). α-1,6 –fucosylation and the fucosyltransferase responsible for this modification are related to several physiological and pathological issues, e.g. it plays an important role in retinal reaggregation (Stelck et al, 1999) and the enzymatic activity and protein expression of FUT8 are increased in tumor tissues of human colorectal carcinoma (Muinelo-Romay et al, 2008). Little information is available of α-1,6 –fucosylation in snails (Gutternigg et al, 2004; Gutternigg et al, 2007). Therefore, we wanted to identify, express and characterize one from gastropods.

None of the species we were working with (*A. lusitanicus*, *A. fulica*, *B. glabrata*) has been sequenced far enough while the project was ongoing. Some sequences were available from the genome project from *B. glabrata* but none of them were helpful for our aims. In order to get the enzyme, three approaches were tried:

- expression library and screening for enzyme activity as described by Larsen (Larsen et al, 1989);
- enrichment of microsomal tissue and further purification;
- through homology search by screening the cDNA with degenerate primers.

cDNA from *A. lusitanicus* was successfully cloned in three expression vectors: one for *P. pastoris* (pGAPZ B) and two for insect cells (pBakPAK8 and pENTR1a). The *P. pastoris* cells library consisted of $1,5*10^5$ clones and supposing that no genes were lost during transformation $2,25*10^4$ of them contained individual genes. We expect that *A. lusitanicus* has approximately $2*10^4$ genes, since for example the recently sequenced oyster contains 28027 (Zhang et al, 2012). The ideal size of the library should be approximately $2*10^5$ (Sambrook et al, 1989), therefore this library was not big enough to find low expressed genes. In addition, the library has two major problems: first, the vector could not be linearized since the sequence of the target gene was not known and therefore the library was not integrated into the genome of *P. pastoris* cells. The plasmid was very instable and was lost easily, making expression screening impossible. Second, *P. pastoris* cells do not synthesize the correct acceptor for FUT8 and therefore it was not possible to screen for fucoslyated glycans with lectins. Using a six or eight-base pair restriction enzyme it should be possible to linearize the library without losing too many genes. Nevertheless, the library is available and can be used for screening for other interesting enzymes. In addition, by now there are some glycoengineered *P. pastoris* strains, in which the library could be cloned and used for screening for at least some other glycosyltransferases.

Two different vectors were chosen for the insect cells library: pBacPAK8, which is a well-established expression vector for insect cells, where recombination into the baculovirus genome is done *in vivo* and pENTR1a, where recombination is done *in vitro*. This strategy should be more efficient than the *in vivo* one and therefore improve the size and diversity of the library. The size of the pENTR1a library in *E. coli* cells was very small, just 7×10^2 clones. Many difficulties arose already when linearizing the vector by inverted PCR, as it contains the repetitive aTTR sites which are necessary for recombination and the polymerase seemed to have some difficulties with it. Once the cDNA was cloned, it was never possible to have enough clones to create a library and the few ones always had an insert of the same length. We suppose that *E. coli* cells might have some problems with this particular vector and it could be helpful to directly recombine the vector with the baculovirus genome.

The pBacPAK8 library was not as large as the library generated in *P. pastoris* cells since only 2×10^4 clones were counted but it had high diversity: 33% of the clones had individual genes, which corresponded to $6,6 \times 10^3$ different genes. It was also possible to identify some interesting ORFs, like one of a C-type lectin. Given the restricted size and diversity of the library, and the fact that no normalization was carried out in order to enrich low abundance genes, this approach was not successful. Our method of screening was also based on

screening with fucose-specific lectins, like *Lens culinaris* agglutinin or *Aleuria aurantia* lectin, for enhanced glycosylation. The two major problems with this strategy were that all insect cells have native α-1,6-fucosylated glycans (Palmberger et al, 2011). *Lens culinaris* agglutinin is described in literature to bind to α-mannose structures and alpha-1,6-fucosylated glycans (Gabius et al, 2004). More detailed studies on modified glycoproteins showed that the strongest binding is on $GnGnF^6$ oligosaccharides (for the structure see chapter 10) but it also binds to some extent to GnGn and $GnGnF^3$ oligosaccharides (Iskratsch et al, 2009). The same study showed also that *Aleuria aurantia* lectin not only binds to GalFGalF (for the structure see chapter 10) and $GnGnF^6$ oligosaccahrides but also to some extent to $GnGnF^3$ oligosaccharides. There are also two other lectins with fucose specificity: *Pisum sativum* agglutinin and *Aspergillus oryzae* lectin. The specificity of all the four mentioned lectins was analyzed in (Tateno et al, 2009) with the conclusion that *Pisum sativum* agglutinin has similar binding properties as *Lens culinaris* agglutinin whereas *Aspergillus oryzae* lectin has broad specificity to fucosylated glycans similarly to that *Aleuria aurantia* lectin. These data may explain the completely different behaviour of *Lens culinaris* agglutinin and *Aleuria aurantia* lectin on the first FACS experiment using our library. Considering all these aspects, a proof of concept was also made by cloning the mFUT8 in insect cells. A clear difference in immunofluorescence intensity was visible in the confocal microscopy images and in the FACS experiment between the cells expressing the control protein and the FUT8 expressing ones. The fault in the proof of concept was that all cells were overexpressing the fucosyltransferases which is not the case for the library, where just some cells may contain the correct gene. Our method of screening was not sensitive enough for this case. A very small enhancement in the population with higher fluorescence intensity which could be seen after three rounds of sorting was just a statistical fluctuation. The PCR and FACS control after the three rounds confirmed this. In order to isolate an α-1,6-fucosyltransferase with this strategy, it would be better to have first a bigger library and then transform it in a cell line lacking FUT8 because of a knock out. Instead of FACS, a panning technique as described in (Larsen et al, 1989) would be an additional option to isolate cells with a functional recombinant FUT8.

The second strategy was to go the other way round and purify the enzyme to get a peptide sequence for designing corresponding degenerate primer allowing to screen the cDNA library by PCR. As first step the acceptor substrate for the enzyme activity assay was prepared. The protocol is well established in our lab and we were able to purify a good amount of GnGn oligosaccharide. As an intact GlcNAc at the reducing end is necessary for the FUT8 activity, there is no possibility to label it with 2-AP and further purify it by preparative HPLC. Just an aliquot was labelled in order to check the quality before using the preparation for enzyme activity assay. The quality was satisfying as the main peak was the GnGn oligosaccharide one and just a couple of small peaks mainly from oligosaccharides still containing one or two galactoses were visible.

The GnGn oligosaccharide prepared this way was used in the second step: establishing an enzyme activity assay using HPLC. For this scope labelling was performed after the assay. When measuring FUT8 activity of the homogenised rabbit brain with radioactivity it was visible, but it was not when the sample was applied to HPLC. It seemed not to be easy to establish a non-radioactive assay for FUT8 even if in the past some alternative labelling methods where published, such as labelling GnGn glycopeptides with dansyl-chlorid (Roitinger et al, 1998) or with 4-(2-Pyridylamino) Butylamine (Uozumi et al, 1996) on the amino acids. Both assays work well but the preparation is not easy. In case of the oligopeptide-GnGn labelled with dansyl-chlorid the biggest difficulty is to have homogenous oligopeptides with the same amino acids. On the other hand the with 4-(2-Pyridylamino) Butylamine labelling requires several complicate synthesis steps which makes the whole procedure very laborious.

The next step was to enrich the gastropod tissue for FUT8. The fucosyltransferase is a type II transmembrane protein which is localized in the Golgi. In the past, especially *from Lymnea stagnalis* some glycosyltransferases were characterized this way (Mulder et al, 1995a; Mulder et al, 1991; Mulder et al, 1996; Mulder et al, 1995b). A microsome preparation should

help to enrich the enzyme and get rid of some disturbing factors which are present in the whole tissue homogenate of the snails and slugs to make some first activity measurements. Microsome preparations from *A. lusitanicus* and Bge cells were performed. The assay of the *A. lusitanicus* preparation was analyzed by HPLC and as in case of the rabbit brain homogenate no activity was visible. In this case, there were two uncertainty factors: the enzyme activity assay was still not very well established and the same was true for the microsome preparation. For example, microsome preparations from Bge cells always showed enhanced activity as compared to the controls when measuring them with radioactivity, but the results have to be taken with care as the standard deviations were always very high. Optimization of the microsome preparation or enrichment of tissue with high fucosyltransferases expression could lead to a first characterization of the enzyme.

The third and last strategy used was in fact the most common one: homology search to design degenerate primers for screening the cDNA library. The premises were that sequences from different organism of the searched proteins were available and at least some regions of it were well conserved through the species. This was the case for FUT8 which is very well conserved even at nucleotide level. Nevertheless, many different primer pairs and PCRs were tried until the first sequence piece of FUT8 was found. As many other glycosyltransferases, FUT8 is not highly expressed and the amount in the cDNA library was very low. A nested PCR approach was necessary to amplify it from the *A. lusitanicus* cDNA library and it also helped to amplify the 5' end of the ORF. After these two steps we had 1609 bp corresponding to a 536 amino acids long sequence of approximately 60 kDa. All the characteristic domains of the FUT8 family were visible: the short cytoplasmatic tail, a transmembrane domain, the GDP-fucose binding domain, the catalytic domain and the beginning of the SH3 domain. The last piece of the SH3 domain and the C-terminus were missing and a whole variety of methods were tried to find it. Even if a lot of bands of the correct size were amplified by 3'RACE none of them corresponded to the 3' end of FUT8. Often a splicing variant was found, which is nothing new since in human four variants were reported, one of them being expressed in the retina (Yamaguchi et al, 2000). Inverse PCR on the blunt or sticky end ligated library gave similar unsatisfying result and also southern and colony blot of the cDNA library were unsuccessful. As a last instance we wanted to enrich the library for FUT8 through biotinylated probes and sequence it with a next-generation sequencing method but –probably due to a not well established hybridization protocol for the probes- this was not successful too. For all these methods it would have been helpful to have a bigger normalized library with a higher gene diversity and it would also make sense to analyze the RNA by Northern blot and real-time PCR to check how many splicing variants are present in *A. lusitanicus* and how much they are expressed.

We are still not sure if the segment that we amplified is an active variant or not. A soluble form of the incomplete *A. lusitanicus* FUT8 without the cytoplasmatic tail and the transmembrane domain was successfully expressed in *P. pastoris* and insect cells. Both expression vectors had a secretion signal, but whereas in *P. pastoris* cells some small amount of FUT8 seemed to be in the supernatant, this was not the case in insect cells where the enzyme was just located intracellular. This was not a real problem as after cell lysis the enzyme was in the soluble part but both, the *P. pastoris* and the insect cells expressed enzyme were not active. There are many reasons which could explain this. The first and more obvious one is that the lacking of the SH3 domain and the C-terminus made the protein inactive, as the SH3 may be important for activity, even if until now the exact role is not known (Ihara et al, 2007). An incomplete or modified C-terminus on a glycosyltransferase is also very critical as it often influences the structure and therefore the activity of the protein. As already mentioned, the FUT8 isoform that we isolated from *A. lusitanicus* could also be an inactive one, even if all domains typical for an active form were present. It seems very improbable that *P. pastoris* and insect cells modify the slug enzyme in a way that it was not active as both expression systems were commonly used for glycosyltransferases (Bencúrová et al, 2003) and (Shinkai et al, 1997). There is also a remote possibility that the enzyme had eventual special requirements that we did not know.

In order to prove that the missing activity is due to the lacking of the SH3 domain we built four different fusion proteins using the known soluble part of the *A. lusitanicus* FUT8 and replacing the whole SH3 domain with the SH3 domain of *Crassostrea gigas*, *Lottia gigantea*, *Caenorhabditis elegans* and *Drosophila melanogaster* respectively. The first two were the evolutionary nearest one to the slug but the function of their enzymes has until now not been proved. The last two are the evolutionary nearest organism were the activity of FUT8 was confirmed (Paschinger et al, 2005b). Only the cloning of the *C. gigas* fusion protein was successful since there were some cloning difficulties for the other three. Again, the localization in insect cells was intracellular but the enzyme was in the soluble part, which strongly suggests that it was not missfolded. Nevertheless, it was not active but as the activity of the whole *C. gigas* FUT8 has still not been confirmed it is not clear if this is due to the the *A. lusitanicus* FUT8 or to the *C. gigas* SH3 domain. It would be very helpful to have the other constructs and check their activity.

In order to have a look, if the two new mollusc FUT8 sequences from *Lottia gigantea* and *Crassostrea gigas* corresponded to active enzymes, we expressed both proteins in insect cells. Differently from the *A. lusitanicus* one, it was not possible to identify a transmembrane domain using the prediction softwares available on the web. This is quite strange as all FUT8 known till now are typical type 2 Golgi transmembrane proteins. A remarkable fact is also that the *L. gigantea* sequence is very short (just 471 amino acids) compared to other FUT8s (around 560 amino acids). The whole sequences where cloned and expressed in *Trichoplusia ni* High Five insect cells. The localization of both proteins was again intracellular, in case of the *L. gigantea* in the soluble part and for *C. gigas* it was found in the soluble and the insoluble part. In both cases it was a very strange behaviour which raises questions on the structure and functionality of these enzymes. Further structure studies and maybe purification could help to get deeper insight on these molluscan FUT8s since till now we could not confirm any activity. It is very interesting that they behaved exactly as *A. lusitanicus* FUT8, which could be an indication that molluscan FUT8s may be different and have other requirements for their activity.

As a side project, purification of a ß- galactosidase from *A. lusitanicus* was performed. β-galactosidase is an exoglycosidase which plays a major role in glycan modification and has been shown to work on secreted glycoproteins by leading to a change on the heterogeneity of glycoforms. The purification protocol was pretty well established but a real purification of the enzyme was never achieved and the peptide sequence was still unknown. This time we wanted to have enough material to get the peptide sequence and design degenerate primer to amplify the whole sequence from the cDNA library. The major issue during purification was that some ß- galactosidase can still be seen in the first fractions of substrate analogue affinity chromatography even after three rounds of purification. Probably there were more ß-galactosidase isoforms which had different substrate specificities and therefore did not bind to the column. In *Achatina achatina*, for example, the existence of one specific and two unspecific galactosidases has been reported (Leparoux et al, 1997). It is also nearly impossible to get enough material to analyze from one purification only, as scale-up often led to column failure. It would be helpful to start with a much purer material, such as the digestive juice as described in (Leparoux et al, 1997). Nevertheless, by putting different purifications together there was enough material to get a couple of peptide sequences but they did not match to any glycosidases in the databases. Maybe a screening of the cDNA libraries using degenerate primers designed by aligning the sequences of the recently partially or completely sequenced molluscs such as *C. gigas*, *L. gigantea* or *A. californica* could be successful.

6 References

Aeed PA, Elhammer AP (1994) Glycosylation of recombinant prorenin in insect cells: the insect cell line Sf9 does not express the mannose 6-phosphate recognition signal. *Biochemistry* **33:** 8793-8797

Bakker H, Agterberg M, Van Tetering A, Koeleman CA, Van den Eijnden DH, Van Die I (1994) A Lymnaea stagnalis gene, with sequence similarity to that of mammalian beta 1-->4-galactosyltransferases, encodes a novel UDP-GlcNAc:GlcNAc beta-R beta 1-->4-N-acetylglucosaminyltransferase. *J Biol Chem* **269:** 30326-30333

Bakker H, Schoenmakers PS, Koeleman CA, Joziasse DH, van Die I, van den Eijnden DH (1997) The substrate specificity of the snail Lymnaea stagnalis UDP-GlcNAc:GlcNAc beta-R beta 4-N-acetylglucosaminyltransferase reveals a novel variant pathway of complex-type oligosaccharide synthesis. *Glycobiology* **7:** 539-548

Bencúrová M, Rendić D, Fabini G, Kopecky EM, Altmann F, Wilson IB (2003) Expression of eukaryotic glycosyltransferases in the yeast Pichia pastoris. *Biochimie* **85:** 413-422

Bretting H, Messer M, Bornaghi L, Kröger L, Mischnick P, Thiem J (2000) Galactan biosynthesis in snails: a comparative study of beta-(1--> 6) galactosyltransferases from Helix pomatia and Biomphalaria glabrata. *J Comp Physiol B* **170:** 601-613

Brockhausen I, Schachter H, Stanley P (2009) O-GalNAc Glycans. In *Essentials of Glycobiology*, Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. The Consortium of Glycobiology Editors, La Jolla, California

Brzezinski K, Rogozinski B, Stepkowski T, Bujacz G, Jaskolski M (2004) Cloning, purification, crystallization and preliminary crystallographic studies of Bradyrhizobium fucosyltransferase NodZ. *Acta Crystallographica Section D* **60:** 344-346

Bürgmayr S, Grabher-Meier H, Staudacher E (2001) Sialic acids in gastropods. *FEBS Lett* **508:** 95-98

Cabezas JA, Reglero A, Calvo P (1983) Glycosidases. (Fucosidases, galactosidases, glucosidases, hexosaminidases and glucuronidase from some molluscs and vertebrates, and neuraminidase from virus). *International Journal of Biochemistry* **15:** 243-259

Calvo P, Santamaria MG, Melgar MJ, Cabezas JA (1983) Kinetic evidence for two active sites in β-d-fucosidase of Helicella ericetorum. *International Journal of Biochemistry* **15:** 685-693

Colas B (1980) Kinetic studies on β-fucosidases of Achatina balteata. *BBA - Enzymology* **613:** 448-458

Colas B, Attias J (1977) Purification of two β (D) glycosidases from the digestive juice of Achatina balteata. *PURIFICATION DE DEUX β D GLYCOSIDASES DU SUC DIGESTIF D'ACHATINA BALTEATA* **69:** 577-585

Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* **328:** 307-317

Dijk J, Brouwer M, Coert A, Gruber M (1970) Structure and function of hemocyanins. VII. The smallest subunit of alpha- and beta-hemocyanin of Helix pomatia: size, composition, N- and C-terminal amino acids. *Biochim Biophys Acta* **221:** 467-479

Dimri GP, Lee X, Basile G, Acosta M, Scott G, Roskelley C, Medrano EE, Linskens M, Rubelj I, Pereira-Smith O, et al. (1995) A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proc Natl Acad Sci U S A* **92:** 9363-9367

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

Durocher Y, Butler M (2009) Expression systems for therapeutic glycoprotein production. *Current Opinion in Biotechnology* **20:** 700-707

Elbein AD (1987) Inhibitors of the biosynthesis and processing of N-linked oligosaccharide chains. *Annu Rev Biochem* **56:** 497-534

Ellerton HD, Ellerton NF, Robinson HA (1983) Hemocyanin--a current perspective. *Prog Biophys Mol Biol* **41:** 143-248

Ernst LK, Rajan VP, Larsen RD, Ruff MM, Lowe JB (1989) Stable expression of blood group H determinants and GDP-L-fucose: beta-D-galactoside 2-alpha-L-fucosyltransferase in mouse cells after transfection with human DNA. *J Biol Chem* **264:** 3436-3447

Freeze HH, Schachter H (2009) Genetic Disorders of Glycosylation. In *Essentials of Glycobiology*, Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. The Consortium of Glycobiology Editors, La Jolla, California

Frýda J, Hausdorf B, Ponder W, Valdes A, Warén A (2005) Classification and nomenclator of gastropod families. In *Malacologia: International Journal of Malacology*, Bouchet P, Rocroi J-P (eds), Vol. 47. Hackenheim, Germany: ConchBooks

Fuhrmann U, Bause E, Ploegh H (1985) Inhibitors of oligosaccharide processing. *Biochim Biophys Acta* **825:** 95-110

Furukawa K, Tsuchida A, Okajima T (2009) Glycoconjugate glycosyltransferases. *Glycoconj J* **26:** 987-998

Gabius HJ, Siebert HC, Andre S, Jimenez-Barbero J, Rudiger H (2004) Chemical biology of the sugar code. *Chembiochem* **5:** 740-764

Geyer H, Wuhrer M, Kurokawa T, Geyer R (2004) Characterization of keyhole limpet hemocyanin (KLH) glycans sharing a carbohydrate epitope with Schistosoma mansoni glycoconjugates. *Micron* **35:** 105-106

Gielens C, Idakieva K, Van den Bergh V, Siddiqui NI, Parvanova K, Compernolle F (2005) Mass spectral evidence for N-glycans with branching on fucose in a molluscan hemocyanin. *Biochemical and Biophysical Research Communications* **331:** 562-570

Got R, Marnay A (1968) Isolation, purification and some physicochemical characteristics of 2 beta-hexosidases of the digestive juice of Helix pomatia. *Isolement, purification et quelques caractéristiques physicochimiques de deux beta-hexosidases du suc digestif d'Helix pomatia* **4:** 240-246

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

Hall RL, Wood EJ (1976) The carbohydrate content of gastropod haemocyanins. *Biochem Soc Trans* **4:** 307-309

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 E (1976) A cell line from embryos of *Biomphalaria glabrata* (Pulmonata): Establishment and characteristics. In *Invertebrate tissue culture: research applications*, Maramorosch K (ed). New York: Academic Press

Hansen SF, Bettler E, Rinnan A, Engelsen SB, Breton C (2010) Exploring genomes for glycosyltransferases. *Mol Biosyst* **6:** 1773-1781

Helenius A (1994) How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol Biol Cell* **5:** 253-265

Helenius A (1997) Calnexin, calreticulin and the folding of glycoproteins. *Trends in cell biology* **7:** 193-200

Helenius J, Ng DTW, Marolda CL, Walter P, Valvano MA, Aebi M (2002) Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* **415:** 447-450

Hokke CH, Deelder AM (2001) Schistosome glycoconjugates in host-parasite interplay. *Glycoconj J* **18:** 573-587

Hutchinson WL, Johnson PJ, Du MQ, Williams R (1991) Serum and tissue alpha-L-fucosidase activity in the pre-clinical and clinical stages of hepatocellular carcinoma. *Clin Sci (Lond)* **81:** 177-182

Ihara H, Ikeda Y, Toma S, Wang X, Suzuki T, Gu J, Miyoshi E, Tsukihara T, Honke K, Matsumoto A, Nakagawa A, Taniguchi N (2007) Crystal structure of mammalian alpha1,6-fucosyltransferase, FUT8. *Glycobiology* **17:** 455-466

Iskratsch T, Braun A, Paschinger K, Wilson IB (2009) Specificity analysis of lectins and antibodies using remodeled glycoproteins. *Anal Biochem* **386:** 133-146

Ito Y, Miyauchi A, Yoshida H, Uruno T, Nakano K, Takamura Y, Miya A, Kobayashi K, Yokozawa T, Matsuzuka F, Taniguchi N, Matsuura N, Kuma K, Miyoshi E (2003) Expression of alpha1,6-fucosyltransferase (FUT8) in papillary carcinoma of the thyroid: its linkage to biological aggressiveness and anaplastic transformation. *Cancer Lett* **200:** 167-172

Jacobs PP, Callewaert N (2009) N-glycosylation engineering of biopharmaceutical expression systems. *Curr Mol Med* **9:** 774-800

Khoo K-H, Chatterjee D, Caulfield JP, Morris HR, Dell A (1997) Structural characterization of glycosphingolipids from the eggs of Schistosoma mansoni and Schistosoma japonicum. *Glycobiology* **7:** 653-661

Knight M, Ittiprasert W, Odoemelam EC, Adema CM, Miller A, Raghavan N, Bridger JM (2011) 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

Kobayashi T, Shinnoh N, Goto I, Kuroiwa Y (1985) Hydrolysis of galactosylceramide is catalyzed by two genetically distinct acid beta-galactosidases. *J Biol Chem* **260:** 14982-14987

Kojima N, Tachida Y, Yoshida Y, Tsuji S (1996) Characterization of mouse ST8Sia II (STX) as a neural cell adhesion molecule-specific polysialic acid synthase. Requirement of core α1,6- linked fucose and a polypeptide chain for polysialylation. *Journal of Biological Chemistry* **271:** 19457-19463

Kornfeld R, Kornfeld S (1985) Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* **54:** 631-664

Kurokawa T, Wuhrer M, Lochnit G, Geyer H, Markl J, Geyer R (2002) Hemocyanin from the keyhole limpet Megathura crenulata (KLH) carries a novel type of N-glycans with Gal(beta1-6)Man-motifs. *Eur J Biochem* **269:** 5459-5473

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680-685

Langdon RH, Cuccui J, Wren BW (2009) N-linked glycosylation in bacteria: an unexpected application. *Future Microbiol* **4:** 401-412

Larsen RD, Rajan VP, Ruff MM, Kukowska-Latallo J, Cummings RD, Lowe JB (1989) Isolation of a cDNA encoding a murine UDPgalactose:beta-D-galactosyl- 1,4-N-acetyl-D-glucosaminide alpha-1,3-galactosyltransferase: expression cloning by gene transfer. *Proc Natl Acad Sci U S A* **86:** 8227-8231

Lee BY, Han JA, Im JS, Morrone A, Johung K, Goodwin EC, Kleijer WJ, DiMaio D, Hwang ES (2006) Senescence-associated beta-galactosidase is lysosomal beta-galactosidase. *Aging Cell* **5:** 187-195

Lehr T, Geyer H, Maass K, Doenhoff MJ, Geyer R (2007) Structural characterization of N-glycans from the freshwater snail Biomphalaria glabrata cross-reacting with Schistosoma mansoni glycoconjugates. *Glycobiology* **17:** 82-103

Leparoux S, Padrines M, Placier G, Colas B (1997) Characterization of a strictly specific acid β-galactosidase from Achatina achatina. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1336:** 522-532

Li J, Jänne PA, Makrigiorgos GM (2011) Biotinylated probe isolation of targeted gene region improves detection of T790M epidermal growth factor receptor mutation via peptide nucleic acid-enriched real-time PCR. *Clin Chem* **57:** 770-773

Lommel M, Strahl S (2009) Protein O-mannosylation: conserved from bacteria to humans. *Glycobiology* **19:** 816-828

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

Longmore GD, Schachter H (1982) Product-identification and substrate-specificity studies of the GDP-L-fucose:2-acetamido-2-deoxy-beta-D-glucoside (FUC goes to Asn-linked GlcNAc) 6-alpha-L-fucosyltransferase in a Golgi-rich fraction from porcine liver. *Carbohydr Res* **100:** 365-392

Loos A, Steinkellner H (2012) IgG-Fc glycoengineering in non-mammalian expression hosts. *Arch Biochem Biophys* **526:** 167-173

Luo Y, Haltiwanger RS (2005) O-fucosylation of notch occurs in the endoplasmic reticulum. *J Biol Chem* **280:** 11289-11294

Lõoke M, Kristjuhan K, Kristjuhan A (2011) Extraction of genomic DNA from yeasts for PCR-based applications. *Biotechniques* **50:** 325-328

Lüttge H, Heidelberg T, Stangier K, Thiem J, Bretting H (1997) The specificity of an α-(1 → 2)-l-galactosyltransferase from albumen glands of the snail Helix pomatia. *Carbohydrate Research* **297:** 281-288

Ma B, Simala-Grant JL, Taylor DE (2006) Fucosylation in prokaryotes and eukaryotes. *Glycobiology* **16:** 158R-184R

McArthur AG, Harasewych MG (2003) Molecular systematics of the major lineages of the Gastropoda. In *Molecular Systematics and Phylogeography of Mollusks*, Lydeard C, Lindberg DR (eds), pp 140-160. Washington: Smithsonian Books

Muinelo-Romay L, Vázquez-Martín C, Villar-Portela S, Cuevas E, Gil-Martín E, Fernández-Briera A (2008) Expression and enzyme activity of alpha(1,6)fucosyltransferase in human colorectal cancer. *Int J Cancer* **123:** 641-646

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

Noronkoski T, Mononen I (1997) Influence of L-fucose attached α1→6 to the asparagine-linked N-acetylglucosamine on the hydrolysis of the N-glycosidic linkage by human glycosylasparaginase. *Glycobiology* **7:** 217-220

Okuyama N, Ide Y, Nakano M, Nakagawa T, Yamanaka K, Moriwaki K, Murata K, Ohigashi H, Yokoyama S, Eguchi H, Ishikawa O, Ito T, Kato M, Kasahara A, Kawano S, Gu J, Taniguchi N, Miyoshi E (2006) Fucosylated haptoglobin is a novel marker for pancreatic cancer: a detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation. *Int J Cancer* **118:** 2803-2808

Oriol R, Mollicone R, Cailleau A, Balanzino L, Breton C (1999) Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria. *Glycobiology* **9:** 323-334

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

Paschinger K, Fabini G, Schuster D, Rendić D, Wilson IB (2005a) Definition of immunogenic carbohydrate epitopes. *Acta Biochim Pol* **52:** 629-632

Paschinger K, Rendic D, Lochnit G, Jantsch V, Wilson IB (2004) Molecular basis of anti-horseradish peroxidase staining in Caenorhabditis elegans. *J Biol Chem* **279:** 49588-49598

Paschinger K, Staudacher E, Stemmer U, Fabini G, Wilson IB (2005b) Fucosyltransferase substrate specificity and the order of fucosylation in invertebrates. *Glycobiology* **15:** 463-474

Prieto PA, Larsen RD, Cho M, Rivera HN, Shilatifard A, Lowe JB, Cummings RD, Smith DF (1997) Expression of human H-type alpha1,2-fucosyltransferase encoding for blood group H(O) antigen in Chinese hamster ovary cells. Evidence for preferential fucosylation and truncation of polylactosamine sequences. *J Biol Chem* **272:** 2089-2097

Qasba PK, Ramakrishnan B, Boeggeman E (2005) Substrate-induced conformational changes in glycosyltransferases. *Trends in biochemical sciences* **30:** 53-62

Rademacher TW, Parekh RB, Dwek RA (1988) Glycobiology. *Annu Rev Biochem* **57:** 785-838

Rini J, Esko J, Varki A (2009) Glycosyltransferases and Glycan-processing Enzymes. In *Essentials of Glycobiology*, Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. The Consortium of Glycobiology Editors, La Jolla, California

Roitinger A, Leiter H, Staudacher E, Altmann F (1998) HPLC method for the determination of Fuc to Asn-linked GlcNAc fucosyltransferases. *Glycoconj J* **15:** 89-91

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Sandra K, Dolashka-Angelova P, Devreese B, Van Beeumen J (2007) New insights in Rapana venosa hemocyanin N-glycosylation resulting from on-line mass spectrometric analyses. *Glycobiology* **17:** 141-156

Schachter H (2000) The joys of HexNAc. The synthesis and function of N- and O-glycan branches. *Glycoconj J* **17:** 465-483

Schnaar RL, Suzuki A, Stanley P (2009) Glycosphingolipids. In *Essentials of Glycobiology*, Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. The Consortium of Glycobiology Editors, La Jolla, California

Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Meng YG, Weikert SHA, Presta LG (2002) Lack of Fucose on Human IgG1 N-Linked Oligosaccharide Improves Binding to Human FcγRIII and Antibody-dependent Cellular Toxicity. *J Biol Chem* **277:** 26733-26740

Shinkai A, Shinoda K, Sasaki K, Morishita Y, Nishi T, Matsuda Y, Takahashi I, Anazawa H (1997) High-level expression and purification of a recombinant human alpha-1, 3-fucosyltransferase in baculovirus-infected insect cells. *Protein Expr Purif* **10:** 379-385

Shinkawa T, Nakamura K, Yamane N, Shoji-Hosaka E, Kanda Y, Sakurada M, Uchida K, Anazawa H, Satoh M, Yamasaki M, Hanai N, Shitara K (2003) The Absence of Fucose but Not the Presence of Galactose or Bisecting N-Acetylglucosamine of Human IgG1 Complex-type Oligosaccharides Shows the Critical Role of Enhancing Antibody-dependent Cellular Cytotoxicity. *J Biol Chem* **278:** 3466-3473

Sinnott ML (1990) Catalytic mechanism of enzymic glycosyl transfer. *Chemical Reviews* **90:** 1171-1202

Spiro RG (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12:** 43R-56R

Srivatsan J, Smith DF, Cummings RD (1992) The human blood fluke Schistosoma mansoni synthesizes glycoproteins containing the Lewis X antigen. *J Biol Chem* **267:** 20196-20203

Stangier K, Lüttge H, Thiem JE, Bretting H (1995) Biosynthesis of the storage polysaccharide from the snail Biomphalaria glabrata, identification and specificity of a branching beta 1-->6 galactosyltransferase. *J Comp Physiol B* **165:** 278-290

Stanley P, Schachter H, Taniguchi N (2009) N-Glycans. In *Essentials of Glycobiology*, Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. The Consortium of Glycobiology Editors, La Jolla, California

Staudacher E (2012) Methylation--an uncommon modification of glycans. *Biol Chem* **393:** 675-685

Staudacher E, Altmann F, Glössl J, März L, Schachter H, Kamerling JP, Hård K, Vliegenthart JF (1991) GDP-fucose: beta-N-acetylglucosamine (Fuc to (Fuc alpha 1----6GlcNAc)-Asn-

peptide)alpha 1----3-fucosyltransferase activity in honeybee (Apis mellifica) venom glands. The difucosylation of asparagine-bound N-acetylglucosamine. *Eur J Biochem* **199:** 745-751

Staudacher E, Altmann F, März L, Hård K, Kamerling JP, Vliegenthart JF (1992) Alpha 1-6(alpha 1-3)-difucosylation of the asparagine-bound N-acetylglucosamine in honeybee venom phospholipase A2. *Glycoconj J* **9:** 82-85

Staudacher E, Altmann F, Wilson IB, März L (1999) Fucose in N-glycans: from plant to man. *Biochim Biophys Acta* **1473:** 216-236

Staudacher E, Dalik T, Wawra P, Altmann F, März L (1995) Functional purification and characterization of a GDP-fucose: beta-N-acetylglucosamine (Fuc to Asn linked GlcNAc) alpha 1,3-fucosyltransferase from mung beans. *Glycoconj J* **12:** 780-786

Staudacher E, Marz L (1998) Strict order of (Fuc to Asn-linked GlcNAc) fucosyltransferases forming core-difucosylated structures. *Glycoconj J* **15:** 355-360

Stelck S, Robitzki A, Willbold E, Layer PG (1999) Fucose in alpha(1-6)-linkage regulates proliferation and histogenesis in reaggregated retinal spheroids of the chick embryo. *Glycobiology* **9:** 1171-1179

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

Struppe E, Staudacher E (2000) Occurence of GDP-L-fucose: beta-N-acetylglucosamine (Fuc to asn-linked GlcNAc) alpha 1,6-fucosyltransferases in porcine, sheep, bovine, rabbit and chicken tissues. *Biochim Biophys Acta* **1475:** 360-368

Summers M, Smith GE, Texas Agricultural Experiment Station. (1987) *A manual of methods for baculovirus vectors and insect cell culture procedures*, College Station, Tex.: Texas Agricultural Experiment Station.

Takahashi M, Kuroki Y, Ohtsubo K, Taniguchi N (2009) Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins. *Carbohydrate Research* **344:** 1387-1390

Takahashi T, Ikeda Y, Miyoshi E, Yaginuma Y, Ishikawa M, Taniguchi N (2000) alpha1,6fucosyltransferase is highly and specifically expressed in human ovarian serous adenocarcinomas. *Int J Cancer* **88:** 914-919

Tanaka H, Suzuki K (1977) Substrate specificities of the two genetically distinct human brain β-galactosidases. *Brain Research* **122:** 325-335

Taniguchi N, Miyoshi E, Jianguo G, Honke K, Matsumoto A (2006) Decoding sugar functions by identifying target glycoproteins. *Current Opinion in Structural Biology* **16:** 561-566

Tateno H, Nakamura-Tsuruta S, Hirabayashi J (2009) Comparative analysis of core-fucose-binding lectins from Lens culinaris and Pisum sativum using frontal affinity chromatography. *Glycobiology* **19:** 527-536

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* **30:** 825-833

Ten Hagen KG, Fritz TA, Tabak LA (2003) All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology* **13:** 1R-16R

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:** 177-183

Tomiya N, Narang S, Lee Y, Betenbaugh M (2004) Comparing N-glycan processing in mammalian cell lines to native and engineered lepidopteran insect cell lines. *Glycoconjugate Journal* **21:** 343-360

Tretter V, Altmann F, Kubelka V, März L, Becker WM (1993) Fucose alpha 1,3-linked to the core region of glycoprotein N-glycans creates an important epitope for IgE from honeybee venom allergic individuals. *Int Arch Allergy Immunol* **102:** 259-266

Uozumi N, Teshima T, Yamamoto T, Nishikawa A, Gao YE, Miyoshi E, Gao CX, Noda K, Islam KN, Ihara Y, Fujii S, Shiba T, Taniguchi N (1996) A fluorescent assay method for GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha 1-6fucosyltransferase activity, involving high performance liquid chromatography. *J Biochem* **120:** 385-392

Vagin O, Kraut JA, Sachs G (2009) Role of N-glycosylation in trafficking of apical membrane proteins in epithelia. *Am J Physiol Renal Physiol* **296:** F459-469

Van den Steen P, Rudd PM, Dwek RA, Opdenakker G (1998) Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol* **33:** 151-208

van Der Wel H, Morris HR, Panico M, Paxton T, North SJ, Dell A, Thomson JM, West CM (2001) A non-Golgi alpha 1,2-fucosyltransferase that modifies Skp1 in the cytoplasm of Dictyostelium. *J Biol Chem* **276:** 33952-33963

Van Kuik JA, Sijbesma RP, Kamerling JP, Vliegenthart JF, Wood EJ (1986) Primary structure of a low-molecular-mass N-linked oligosaccharide from hemocyanin of Lymnaea stagnalis. 3-O-methyl-D-mannose as a constituent of the xylose-containing core structure in an animal glycoprotein. *Eur J Biochem* **160:** 621-625

Van Kuik JA, Sijbesma RP, Kamerling JP, Vliegenthart JF, Wood EJ (1987) Primary structure determination of seven novel N-linked carbohydrate chains derived from hemocyanin of Lymnaea stagnalis. 3-O-methyl-D-galactose and N-acetyl-D-galactosamine as constituents of xylose-containing N-linked oligosaccharides in an animal glycoprotein. *Eur J Biochem* **169:** 399-411

van Kuik JA, van Halbeek H, Kamerling JP, Vliegenthart JF (1985) Primary structure of the low-molecular-weight carbohydrate chains of Helix pomatia alpha-hemocyanin. Xylose as a constituent of N-linked oligosaccharides in an animal glycoprotein. *J Biol Chem* **260:** 13984-13988

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

Varki A (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3:** 97-130

Varki A, Freeze HH, Manzi AE (2009a) Overview of glycoconjugate analysis. *Curr Protoc Protein Sci* **Chapter 12:** Unit 12.11 12.11.11-18

Varki A, Kannagi R, Toole BP (2009b) Glycosylation Changes in Cancer. In *Essentials of Glycobiology*, Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. The Consortium of Glycobiology Editors, La Jolla, California

Varki A, Lowe JB (2009) Biological Roles of Glycans. In *Essentials of Glycobiology*, Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. The Consortium of Glycobiology Editors, La Jolla, California

Vercruysse J, Gabriel S (2005) Immunity to schistosomiasis in animals: an update. *Parasite Immunol* **27:** 289-295

Voynow JA, Kaiser RS, Scanlin TF, Glick MC (1991) Purification and characterization of GDP-L-fucose-N-acetyl beta-D-glucosaminide alpha 1----6fucosyltransferase from cultured human skin fibroblasts. Requirement of a specific biantennary oligosaccharide as substrate. *J Biol Chem* **266:** 21572-21577

Wallenfels K, Weil R (1972) 20 β-Galactosidase. In *The Enzymes*, Paul DB (ed), Vol. Volume 7, pp 617-663. Academic Press

Wang G, Boulton PG, Chan NW, Palcic MM, Taylor DE (1999) Novel Helicobacter pylori alpha1,2-fucosyltransferase, a key enzyme in the synthesis of Lewis antigens. *Microbiology* **145 (Pt 11):** 3245-3253

Wang YM, Hare TR, Won B, Stowell CP, Scanlin TF, Glick MC, Hard K, van Kuik JA, Vliegenthart JF (1990) Additional fucosyl residues on membrane glycoproteins but not a secreted glycoprotein from cystic fibrosis fibroblasts. *Clin Chim Acta* **188:** 193-210

Wildt S, Gerngross TU (2005) The humanization of N-glycosylation pathways in yeast. *Nat Rev Micro* **3:** 119-128

Wilson IB, Harthill JE, Mullin NP, Ashford DA, Altmann F (1998) Core alpha1,3-fucose is a key part of the epitope recognized by antibodies reacting against plant N-linked oligosaccharides and is present in a wide variety of plant extracts. *Glycobiology* **8:** 651-661

Wilson JR, Williams D, Schachter H (1976) The control of glycoprotein synthesis: N-acetylglucosamine linkage to a mannose residue as a signal for the attachment of L-fucose to the asparagine-linked N-acetylglucosamine residue of glycopeptide from alpha1-acid glycoprotein. *Biochem Biophys Res Commun* **72:** 909-916

Wuhrer M, Robijn ML, Koeleman CA, Balog CI, Geyer R, Deelder AM, Hokke CH (2004) A novel Gal(beta1-4)Gal(beta1-4)Fuc(alpha1-6)-core modification attached to the proximal N-acetylglucosamine of keyhole limpet haemocyanin (KLH) N-glycans. *Biochem J* **378:** 625-632

Yamaguchi Y, Ikeda Y, Takahashi T, Ihara H, Tanaka T, Sasho C, Uozumi N, Yanagidani S, Inoue S, Fujii J, Taniguchi N (2000) Genomic structure and promoter analysis of the human alpha1, 6-fucosyltransferase gene (FUT8). *Glycobiology* **10:** 637-643

Yamane-Ohnuki N, Kinoshita S, Inoue-Urakubo M, Kusunoki M, Iida S, Nakano R, Wakitani M, Niwa R, Sakurada M, Uchida K, Shitara K, Satoh M (2004) Establishment of FUT8 knockout Chinese hamster ovary cells: an ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. *Biotechnol Bioeng* **87:** 614-622

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

7 List of figures

Figure 1: Types of protein-linked oligosaccharides. 11

Figure 2: Synthesis of the N-glycan precursor. 13

Figure 3: Synthesis of high mannose, hybrid and complex glycans. 14

Figure 4: Biosynthesis of different mucin – O – glycan core structure. 15

Figure 5: Two N-glycans of the functional unit of *Rapana venosa* 16

Figure 6: Ribbon diagrams of representative GT-A and GT-B folds. 18

Figure 7: The hypothetical model of divergent evolution for the known fucosyltransferase genes. 19

Figure 8: Ds cDNA synthesis from *A. lusitanicus*. 62

Figure 9: PCR screening of *A. lusitanicus* library in pSMART2if 62

Figure 10: PCR screening of *A. lusitanicus* library in pGAPZ B. 63

Figure 11: PCR screening of *A. lusitanicus* library in pBakPAK8 63

Figure 12: PCR screening of *A. lusitanicus* library in pENTR1a 63

Figure 13: pBLAST output of one of the sequenced clones, showing the homology to a slug C-type lectin. 64

Figure 14A and B: Transformed *P. pastoris*, A: with library, B: negative control without library. 64

Figure 15: PCR screening of colonies containing pENTR1a with the mFUT8 sequence 64

Figure 16A and B: FACS plot of cells expressing the control protein (A) and mFUT8 (B).. ...65

Figure 17A and B: Immunofluorescence (A) and brightfield (B) microscopy images of *Sf*9 cells expressing the control protein. 66

Figure 18A and B: Immunofluorescence (A) and brightfield (B) microscopy images of *Sf*9 cells expressing mFUT8. 66

Figure 19A and B: FACS plot of the *Aleuria aurantia* lectin staining. 67

Figure 20A and B: FACS plot of the *Lens culinaris* agglutinin staining 67

Figure 21A and B: FACS plot of the first round of sorting 68

Figure 22A and B: FACS plot of the second round of sorting 68

Figure 23: Summary of the FACS control of the three rounds of sorting 69

Figure 24: PCR of the baculovirus stocks after the three rounds of sorting 69

Figure 25: alignment of the nucleotide sequences of FUT8s from *Homo sapiens* (CAA76986.1), *Mus musculus* (AAH10666.1), *Drosophila melanogaster* (AAF48079.1) and *Caenorhabditis elegans* (CAD54736.1). 70

Figure 26: alignment of the amino acid sequences of FUT8s from *Homo sapiens* (CAA76986.1), *Mus musculus* (AAH10666.1), *Drosophila melanogaster* (AAF48079.1) and *Caenorhabditis elegans* (CAD54736.1) 71

Figure 27: Ds cDNA synthesis from *A. lusitanicus*. 71

Figure 28: Gel of the nested PCR 72

Figure 29: PCR screening of the *A. fulica* and *A. lusitanicus* plate after cloning the insert of the nested PCR. 72

Figure 30: pBLAST output after translating the nucleotide sequence....73

Figure 31: Successful nested 5'RACE PCR. 73

Figure 32: Amino acid sequence of *A. lusitanicus* FUT8. 74

Figure 33: Domains recognized by the cds databas 74

Figure 34: Plot of the transmembrane prediction software 74

Figure 35: Alignment of the splicing variant 75

Figure 36: Section of the colony blot membrane after the colouring reaction 76

Figure 37A, B and C: Southern blot with chemiluminescence detection after 2 (A), 15 (B) and 30 (C) min exposition....76

Figure 39: 280 nm measurement and orcinol spotting of the gel filtration fractions after pronase digestion 78

Figure 40: 280 nm measurement and orcinol spotting of the gel filtration fractions after desyalilation.. 79

Figure 41: 280 nm measurement and orcinol spotting of the gel filtration fractions after PNGase A digestion 79

Figure 42A, B and C: Chromatograms of the isomaltose standard (A), purified GnGn control (B) and new acceptor sample (C). 80

Figure 43A, B and C: Chromatogram of the enzyme activity assay 81

Figure 44A, B and C: Chromatogram of the enzyme activity assay. 82

Figure 45A and B: Enzyme activity of Bge cells, cell pellet, medium and microsomes of two preparations (A and B) 83

Figure 46: FUT8 from *A. lusitanicus* directly amplified from the directional library....84

Figure 47: PCR screening and control PCR of eight *P. pastoris* clones 84

Figure 48A and B: Coomassie (A) and WB using Penta-His antibody (B) of the six *P. pastoris* clones....85

Figure 49A and B: Coomassie (A) and WB using Penta-His antibody (B) of *Trichoplusia ni* High Five and *Sf*9 cells expressing FUT8 and a His-tagged control protein....86

Figure 50: WB using Penta-His antibody of supernatant and pellet of *Ascalapha odorata Ao*38 and *Trichoplusia ni* High Five cells expressing FUT8 87

Figure 51A and B: Coomassie (A) and WB using Penta-His antibody (B) of supernatant and pellet of *Trichoplusia ni* High Five cells expressing FUT8 after 2, 3 and 4 days after infection 87

Figure 52: FUT8 fusion protein 88

Figure 53: PCR screening of the fusion protein containing the known soluble part of *A. lusitanicus* and SH3 domain and C-terminus of *C. gigas* 89

Figure 54A and B: Coomassie (A) and WB using Penta-His antibody (B) of supernatant and pellet of *Trichoplusia ni* High Five cells expressing the *C. gigas* fusion protein....89

Figure 55A and B: Coomassie (A) and WB using Penta-His antibody (B) of soluble and insoluble part after lysis of *Trichoplusia ni* High Five cells expressing the *C. gigas* fusion protein 90

Figure 56: PCR screening of FUT8 from *C. gigas* and *L. gigantea* cloned in pVT-Bac 91

Figure 57A and B: Coomassie (A) and WB using Penta-His antibody (B) of the expression of *Crassostrea gigas* and *Lottia gigantea* FUT8 in *Trichoplusia ni* High Five 91

Figure 58: Protein content (black) and enzyme activity (red) after HIC. 92

Figure 59: Protein content (black) and enzyme activity (red) after DE52.....93

Figure 60: Protein content (black) and the enzyme activity (red) after S200. 94

Figure 61: Activity of different exoglycosidases after S200. 94

Figure 62A and B: Protein content and the enzyme activity of different exoglycosidases after the aminophenyl column (A) and the aminobenzyl column (B). 95

Figure 63A and B: ß-galactosidase enzyme activity of second (A) and third (B) round of affinity chromatography with the aminobenzyl column. 96

Figure 64A and B: Coomassie (A) and silver staining (B) of the aminobenzyl affinity chromatographic step. 97

8 List of tables

Table 1: Transmembrane prediction software.....57

Table 2: Primer sequences for the nested PCR.....71

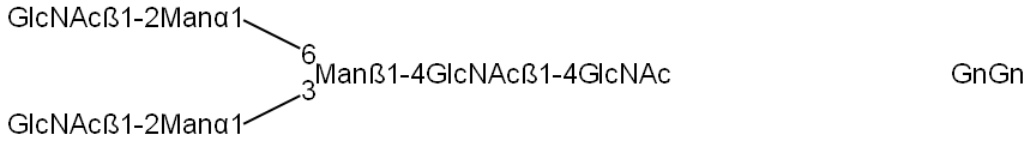
Table 3: Peptide sequences from the ß-galactosidase purification. 97

9 Table of abbreviations

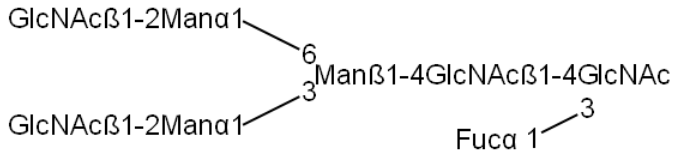
2-AP	2-aminopyrdine
BCIP/NPT	5-bromo-4-chloro-3-indolyl phosphate/nitrobluetetrazolium
Bge	*B. glabrata* embryonic
Biotin-TEG	Biotin with 15 carbon atom spacer
BSA	Bovine serum albumin
CV	Column volume
DIG	Digoxigenin
Dol-P	Dolichol phosphate
Dol-P-P-GlcNAc	Pyrophosphate N-acetylglucosamine
ds cDNA	Double stranded cDNA
EGF	Epidermal growth factor
ER	Endoplasmatic reticulum
FACS	Fluorescence-activated cell sorting
FCS	Fetal calf serum
FUT8	α-1,6-fucosyltransferase
GalNAc	N-acetylgalactosamine
GlcNAc	N-acetylglucosamine
GlcNAcT	N-acetylglucosaminyltransferase
LB	Lysogeny broth
LB LS	LB low salt
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
MES	2-(N-morpholino) ethanesulfonic acid
mFUT8	Mouse α-1,6-fucosyltransferase
ORF	Open reading frame
PBS	Phosphate Buffer Saline
PE	Phycoerythrin
PNGase A	N-glycopeptidase A
pNP	p-Nitrophenol, 1-Hydroxy-4-nitrobenzol
ppGalNAcT	polypeptide N-acetylgalactosamine transferase
RACE	Rapid amplification of cDNA ends
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Ser/Thr	Serine/threonine
*Sf*9	Spodoptera frugiperda *Sf*9 cells
SH3	SRC Homology 3
SSC	Saline-sodium citrate buffer
T4 PNK	T4 polynucleotide kinase
TE buffer	Tris-EDTA Buffer
TGF	Thrombospondin type repeat
WB	Western blot
YPD	Yeast Extract Peptone Dextrose

10 Appendix

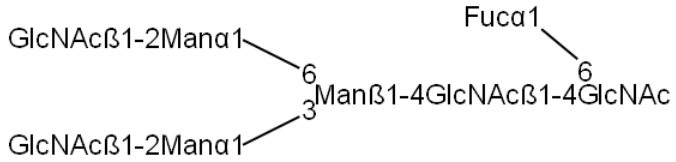
10.1 Glycan structures



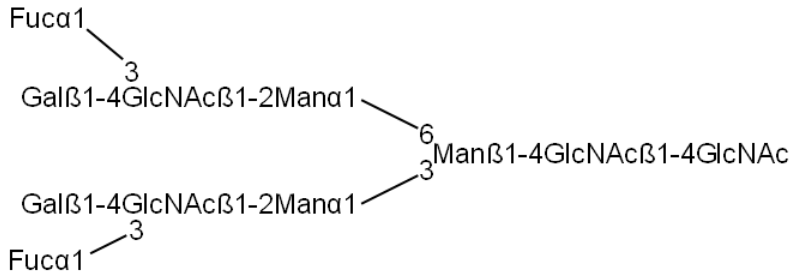
GnGn



GnGnF[3]



GnGnF[6]



GalFGalF