Structure – function study of Arabidopsis thaliana core alpha1,3-fucosyltransferase (FucTA)

Peter Both

UNIVERSITE JOSEPH FOURIER (GRENOBLE I)
ECOLE DOCTORALE CHIMIE ET SCIENCES DU VIVANT

UNIVERSITE SLOVAQUE DE TECHNOLOGIE DE BRATISLAVA (STUBA)

THESE

Pour l'obtention du Diplôme de
DOCTEUR DE L'UNIVERSITE JOSEPH FOURIER

Discipline: Biologie

Présentée et soutenue publiquement le 29 octobre 2009 par

PETER BOTH

ETUDE STRUCTURE-FUNCTION D'UNE FUCOSYLTRANSFERASE (FUCT-A) DE *ARABIDOPSIS THALIANA*

JURY

Rapporteurs : **Prof. Abderrahman MAFTAH, Université de Limoges**
Dr Eva HOSTINOVA, Slovak Academy of Sciences, Bratislava

Examinateurs : **Dr Eva KUTEJOVA, Slovak Academy of Sciences, Bratislava**
Dr Serge PEREZ, ESRF, Grenoble
Dr Jan MUCHA, Slovak Academy of Sciences, Bratislava
Prof. Christelle BRETON, Université Grenoble I

Thèse préparée au CERMAV (Grenoble) et ICHSAS (Bratislava)

I would like to express my gratitude to all those who gave me the possibility to complete this thesis.

I am deeply indebted to my supervisors Dr. Ján Mucha and Prof. Christelle Breton whose help, stimulating suggestions and encouragement helped me in the whole time of research for and writing of this thesis as well as for the opportunity they gave me to work on this interesting subject.
I am also grateful to the former director of CERMAV-CNRS Prof. Serge Pérez and the director of ICh-SAS Dr. Igor Tvaroška DSc. for allowing me to work at these institutions. I also want to thank for the Early-stage Marie Curie Fellowship which covered experimental work during 18 months at CERMAV-CNRS.

Besides my supervisors, I would like to thank Ao.Prof. Iain B.H. Wilson for the continuous support of my PhD study and research, for his patience, motivation, enthusiasm, and immense knowledge.

My sincere thanks also go to Margita Plšková, Dr. Stanislav Kozmon, Dr. Katharina Paschinger, Dr. Dubravko Rendić, Dr. Sara Fasmer Hansen, Ao.Prof. Vladimír Farkaš DSc., Dr. Eva Stratilová, Dr. Imrich Barak DSc., Dr. Jaroslav Klaudiny, and Valerie Chazalet.

I wish to thank Camelia Stînga for helping me get through the difficult times, and for all the emotional support and caring she provided**.**

Lastly, and most importantly, I wish to thank my parents. They raised me, supported me, taught me, and loved me. To them I dedicate this thesis.

Préface

Une fucosylation anormale chez l'homme est observée dans un grand nombre de pathologies sévères (inflammation, rhumatismes, diabète de type II, maladies neuro-dégénératives, cancer, …). Elle est aussi parfois responsable d'allergies alimentaires, au pollen et aux insectes. Les fucosylations sont réalisées par des enzymes appelées fucosyltransférases (FucTs), qui catalysent le transfert de résidus fucose à partir de GDP-L-fucose sur divers accepteurs de type glucidique ou protéique. Quand il s'agit d'un accepteur glucidique, le résidu fucose est transféré sur un galactose en formant une liaison de type α1,2, ou sur une *N*-acétylglucosamine (GlcNAc) en formant une liaison de type α1,3-, α1,4-, ou α1,6. La fucosylation de type « core α1,3 » consiste en l'addition d'un résidu fucose sur le GlcNAc proximal d'une chaîne de *N*-glycanne (résidu GlcNAc directement lié à la chaîne latérale d'une asparagine). Cette « core α1,3-fucosylation » n'est observée que chez les plantes et les insectes et elle est responsable chez l'homme de nombreuses réactions allergiques.

Sur la base des homologies de séquence, les glycosyltransférases sont classées en 90 familles distinctes (appelées GTx). Cette classification qui est continuellement mise à jour est disponible dans la banque de données CAZy (http://www.cazy.org/). Les fucosyltransférases connues à ce jour sont retrouvées dans 5 familles (GT10, 11, 23, 37, 65, 68). La famille GT10 comprend l'ensemble des fucosyltransférases catalysant l'addition d'un fucose en α1,3- ou α1,4- sur un GlcNAc. Chez les eucaryotes, la plupart des FucTs sont localisées dans l'appareil de Golgi. Ce sont des protéines transmembranaires de type II, consistant en une courte région N-terminale cytoplasmique, suivie d'un domaine transmembranaire, d'une région tige et d'un large domaine catalytique globulaire. Récemment, la structure cristallographique d'un des membres de la famille GT10 (une α3-FucT bactérienne) a été résolue. Cette fucosyltransférase adopte un repliement de type GT-B qui est constitué de deux domaines de type Rossmann avec le site catalytique situé à l'interface de ces deux domaines.

L'objectif de cette thèse a été de mieux comprendre les relations séquence-structure-fonction des FucTs de la famille GT10, avec un accent particulier sur les core α3-FucTs responsables de la formation de l'épitope immunogénique (épitope HRP, pour Horse Radish Peroxidase) chez l'homme. L'étude a été initiée par une analyse bioinformatique exhaustive des séquences de la famille GT10 afin d'identifier les différents domaines de l'enzyme, les motifs peptidiques conservés

et les résidus invariants susceptibles de jouer un rôle clé dans la catalyse. Outre la présence de deux motifs conservés bien connus dans cette famille de FucTs, nous avons mis en évidence un nouveau motif dans le domaine catalytique. Ce motif (appelé 1st cluster motif) a été détecté grâce à l'utilisation de la méthode d'analyse des amas hydrophobes (méthode HCA) et sa signature est la suivante : **S**-(N/H/D)-X_{5-9}-**R**-X_6-(L/I)-X_{3-7}-(V/L/I)-X_3-**G**. Il présente 3 résidus invariants et plusieurs positions conservées. Par ailleurs, nous avons trouvé un motif **C**XX**C** conservé dans toutes les séquences de FucTs issues des organismes eucaryotes supérieurs. Outre ces 2 résidus Cys, les FucTs eucaryotes présentent un « pattern » conservé de 4 résidus Cys dont il a été démontré chez certaines FucTs animales qu'elles sont impliquées dans la formation de ponts disulfure. D'autres résidus Cys sont conservés dans les α1,3-fucosyltransférases des invertébrés ainsi que dans quelques core α1,3-fucosyltransferases des plantes. L'analyse comparée des séquences des core α3-FucTs de plantes et d'invertébrés révèle des différences intéressantes dans la région N-terminale du domaine catalytique, notamment la présence d'une séquence d'environs 20 résidus qui n'est présente que dans les séquences de plante. Sachant que les enzymes de plante présentent une spécificité plus stricte au niveau de l'accepteur que les enzymes d'invertébré, cette séquence pourrait être responsable de la reconnaissance du GlcNAc distal dans la structure du substrat accepteur des plantes.

Au cours de cette analyse bioinformatique, il a été mis en évidence une particularité concernant les core α3-FucT. Ces enzymes présentent un domaine supplémentaire hautement conservé en C-terminal, d'environ une centaine de résidus, et de fonction inconnue. Notre équipe s'est donc intéressée plus particulièrement à cette sous-famille d'enzymes, les core α3-FucTs, en utilisant comme modèle d'étude, la fucosyltransférase végétale FucTA d'*Arabidopsis thaliana*. L'enzyme a été exprimée avec succès dans le système *Pichia Pastoris* sous forme soluble dépourvue de la partie CTS (domaines Cytosolique, Transmembranaire et de la région tige (Stem)). Les caractéristiques biochimiques de l'enzyme ont été déterminées et son comportement vis-à-vis des cations divalents caractérisé. Nos résultats expérimentaux ont confirmé que les enzymes de la famille GT10 sont sensibles à la manipulation côté C-terminal et que le domaine N-terminal CTS n'est pas essentiel pour l'activité de l'enzyme et peut être remplacé par d'autres séquences (ex : addition d'un tag). Ainsi, un mutant hybride de FucTA, tronqué de 88 résidus à son extrémité N-terminale, et comportant à la place un marquage N-terminal de type 6xHis+Flag, s'est révélé très actif et a été l'objet de départ pour notre étude structure-fonction. L'enzyme reconnaît le GnGn ou le GnGnF6 comme substrat accepteur et initialise la formation de l'épitope HRP *in vitro*. Le pH

optimal pour l'activité de l'enzyme est de 6,5 et la réaction est optimale à 30°C. L'activité de l'enzyme est fortement stimulée par les ions Mn^{2+} et Mg^{2+}, tandis que les ions de Cu^{2+} et Zn^{2+} ont un faible effet inhibiteur. Pour le substrat accepteur GnGnF6-NST-dansyl, la valeur Km pour le GDP-Fuc est de 0,1mM.

Les observations issues de l'analyse bioinformatique ont été validées expérimentalement à travers des expériences de mutagenèse dirigée ciblant des résidus d'acides aminés clés et par des expériences de troncation des différents domaines identifiés afin de tester leur importance dans la fonction catalytique. Nous avons démontré que le domaine unique C-terminal des core α1,3-fucosyltransférases des plantes est essentiel pour l'activité de FucTA. Nous avons remarqué que la suppression de la région (C388-V501) conduit à une diminution drastique de l'activité (<1%) mais pas à sa complète abolition, ce qui suggère que cette région pourrait être impliquée dans la stabilisation de la structure de l'enzyme et/ou dans d'autres fonctions qui restent à déterminer. Il a été montré que les deux résidus conservés, S218 et R226, sont essentiels pour l'activité de l'enzyme FucTA puisque leur mutation en alanine abolit complètement l'activité. Les mutations de deux autres résidus, Y243 et N219, en alanine conduit également à une diminution significative de l'activité de l'enzyme (> 90%). Dans le modèle 3D proposé, le résidu S253 (situé dans une boucle entre le 1er cluster motif et le motif α1,3 FucT) forme une liaison hydrogène avec la guanine du substrat donneur. Le mutant S253A conserve seulement 35% d'activité.

Enfin, parce que les glycosyltransférases sont elles-mêmes des glycoprotéines, l'influence de l'état de glycosylation de la FucTA a été étudiée. Il avait été démontré que la *N*-glycosylation est importante pour maintenir l'activité maximale de quelques α1,3/4-fucosyltransférases humaines. La FucTA de *A. thaliana* possède quatre sites potentiels de *N*-glycosylation (N64, N337, N420 et N481), dont 3 sont localisés dans le domaine catalytique. Dans *Pichia pastoris,* les expressions hétérologues du mutant hybride FucTA mentionné ci-dessus (ne possédant pas le site N64), ont mis en évidence que l'enzyme est bien glycosylée. Cependant, nous ne savons pas si celle-ci est naturellement glycosylée dans la plante source *A. thaliana*. L'analyse des mutants de *N*-glycosylation de la FucTA met en évidence qu'aucun site n'est à lui seul essentiel à l'activité. Par contre, la mutation de chacun des sites potentiels de *N*-glycosylation conduit à une diminution importante de l'activité (>60%). Nous avons également montré par spectrométrie de masse et Western blotting que FucTA produite sous forme recombinante dans *P. pastoris* présente plusieurs

glycoformes. Les résultats montrent que les trois sites de *N*-glycosylation peuvent être potentiellement glycosylés mais que l'occupation de ces sites est incomplète.

Contents

1. Introduction

1.1. Glycobiology, biological functions of sugars

Glycobiology is a quite large-scale discipline, which deals with molecules that contain carbohydrates, covering the synthesis of sugars (including enzymes involved in their biosynthesis and biodegradation), their properties and biological functions. Glycobiology is the last unconquered frontier of molecular and cell biology. Glycans play an immense role in signal transduction, control of cell differentiation, cell adhesion, and innate immunity (Ioffe E *et al.,* 1997; Wells L *et al.,* 2001; Endo T *and* Toda T, 2003; Hokke CH *and* Yazdanbakhsh M, 2005). Amongst the many functions, glycosylation is involved in the stabilisation of glycoproteins as a protection against proteases and/or as a stabiliser of proper folding as well as a signal itself for the so called "quality control" of proper folding (Molinari M *et al.*, 2005). Another famous example for a fuction of glycoconjugates concerns the blood groups ABO antigens that are determined by oligosaccharidic structures (http://www.bioc.aecom.yu.edu/bgmut/systems_info.php?system=lewis). Until this discovery, blood was thought to be universal which often led to death during inappropriate transfusions (Reid ME *and* Lomas-Francis C, 2004). Furthermore a comparative study showed that human B cells contain a quite specific galactosyltransferase involved in IgG glycosylation. This enzyme from patients with rheumatoid arthritis is less effective because of its lower affinity for UDP-Gal in comparison with the enzyme obtained from healthy individuals (Furukawa K *et al.*, 1990). It was also found that the degalactosylated IgG binds less effectively to the C1p and Fc-receptor. However, no decrease in binding to polyclonal rheumatoid factor and protein A was observed with the degalactosylated IgG (Tsuchiya N *et al.*, 1989). This is an example of how the function of glycoproteins may be modified by the different degrees of maturation of their sugar chains.

Glycoconjugates are not just linked to glycoproteins; also lipids can be modified by sugar.

However, in the past, glycolipids have received less attention than glycoproteins. Glycolipids are involved in cell signalling. The gangliosides (one of the most prominent families of glycolipids) can serve two different roles in cell signalling:

- as ligands for specific receptors
- by participating in the formation of specialised protein/lipid domains (rafts) at the plasma membrane, where they can modulate receptor functions.

The signalling function of gangliosides may be a potential therapeutic target in cancer, diabetes and nerve regeneration. It was shown that disruption of the gene, which is responsible for the first step in the formation of almost all glycosphingolipids is embryonically lethal in knock-out mice, while mice lacking enzymes involved in later steps of glycolipid biosynthesis are viable, but show a range of abnormalities like infertility or neurological defects (Takamiya K *et al.*, 1996; Takamiya K *et al.*, 1998; Furukawa K *et al.*, 2002).

A lot of work was done on mammalian glycosylation pathways, but we often still have a poor understanding of glycosylation in plants and invertebrates. It was thought, that plants and invertebrates produce a very limited range of N- and O-linked glycans in comparison to mammalian systems (Wilson IB, 2002; Bencúrová M *et al.*, 2003). However, plant and invertebrate glycosylation machineries are more complicated than first thought. This has been highlighted by the large number of putative glycosyltransferase genes identified in the completed (or nearly completed) genomes now available for many plant and invertebrate species. However, much work remains to be done to translate the existence of these putative genes into knowledge of the glycans that are actually synthesised by these organisms (Wilson IB, 2002; Bencúrová M *et al.*, 2003).

A good example of one of the roles of O-linked glycosylation is O-fucosylation in *Drosophila* and in mammals. O-linked fucose residues are involved in the regulation of particular cell–cell signalling pathways. By modulating the activity of Notch the Fringe protein carries out a scale of indispensable developmental functions. This modulation is achieved by the Fringe-catalysed addition of *N*-acetylglucosamine residues to fucose residues present on Notch (Moloney DJ *at al.*, 2000; Brückner K *et al.*, 2000; Okajima T *et al.*, 2003; Peterson RE *and* McClay DR, 2005).

1.2. Protein glycosylation

1.2.1. Foreword to protein glycosylation

The vast majority of proteins (most plasma-membrane associated and secretory proteins) for which synthesis is directed to the rough endoplasmic reticulum (ER) are glycoproteins. Glycoproteins contain one or more carbohydrate chains. Some glycosylation reactions occur in the lumen of the ER; others, in the lumina of the *cis*-, *medial*-, or *trans*-Golgi cisternae. Thus the presence of certain carbohydrate residues on proteins provides useful markers for following their movement from the ER and through the Golgi cisternae (Lodish H *et al.*, 2004).

The structures of *N*- and *O*-linked oligosaccharides are very different. For instance, O-linked oligosaccharides are linked to peptides via the hydroxyl group of serine or threonine. In the case of collagens, oligosaccharides are linked to the hydroxyl group of hydroxylysine via galactose (Varki A *et al.*, 1999; Lodish H *et al.*, 2004).

N-linked oligosaccharides are a common feature of eukaryotic glycoproteins. All N-linked carbohydrates are linked via N-acetylglucosamine (GlcNAc) to the amino group of the side chain of asparagine. The consensus sequence of a potential N-glycosylation site is Asn-X- Ser/Thr, where X can be any amino acid except proline. But not all such sequences undergo glycosylation (Lodish H *et al.*, 2004).

A group of enzymes known collectively as the dolichol pathway are responsible for the biosynthesis of the dolichol pyrophosphate-$GlcNAc_2Man_9Glc_3$, which takes place at the membrane of the ER. In this biosynthetic pathway dolichol pyrophosphate, a long-chain (75 – 95 carbon atoms) of polyisoprenoid lipid firmly embedded in the ER membrane, is employed as a carrier (Figure 1). The assembly process proceeds in three stages. First, two N-acetylglucosamine residues and five mannose residues are added to the dolichol phosphate through the action of a number of cytoplasmic enzymes that catalyse monosaccharide transfer from sugar nucleotides. Then, in a remarkable process, this large structure is “flipped” through the ER membrane into the lumen of ER. Finally, additional sugars are added by enzymes in the ER lumen, this time with the use of monosaccharides activated by attachment to dolichol phosphate. This process ends with the formation of a 14-residue oligosaccharide attached to dolichol phosphate. The mechanisms of dolichol pyrophosphate-

$GlcNAc_2Man_9Glc_3$ biosynthesis are highly conserved. The precursor oligosaccharide $GlcNAc_2Man_9Glc_3$ is then transferred to nascent polypeptides, the first step in glycoprotein biosynthesis. The $GlcNAc_2Man_9Glc_3$ transferred to proteins is subsequently modified by the addition and removal of sugars (Burda P *and* Aebi M, 1999).

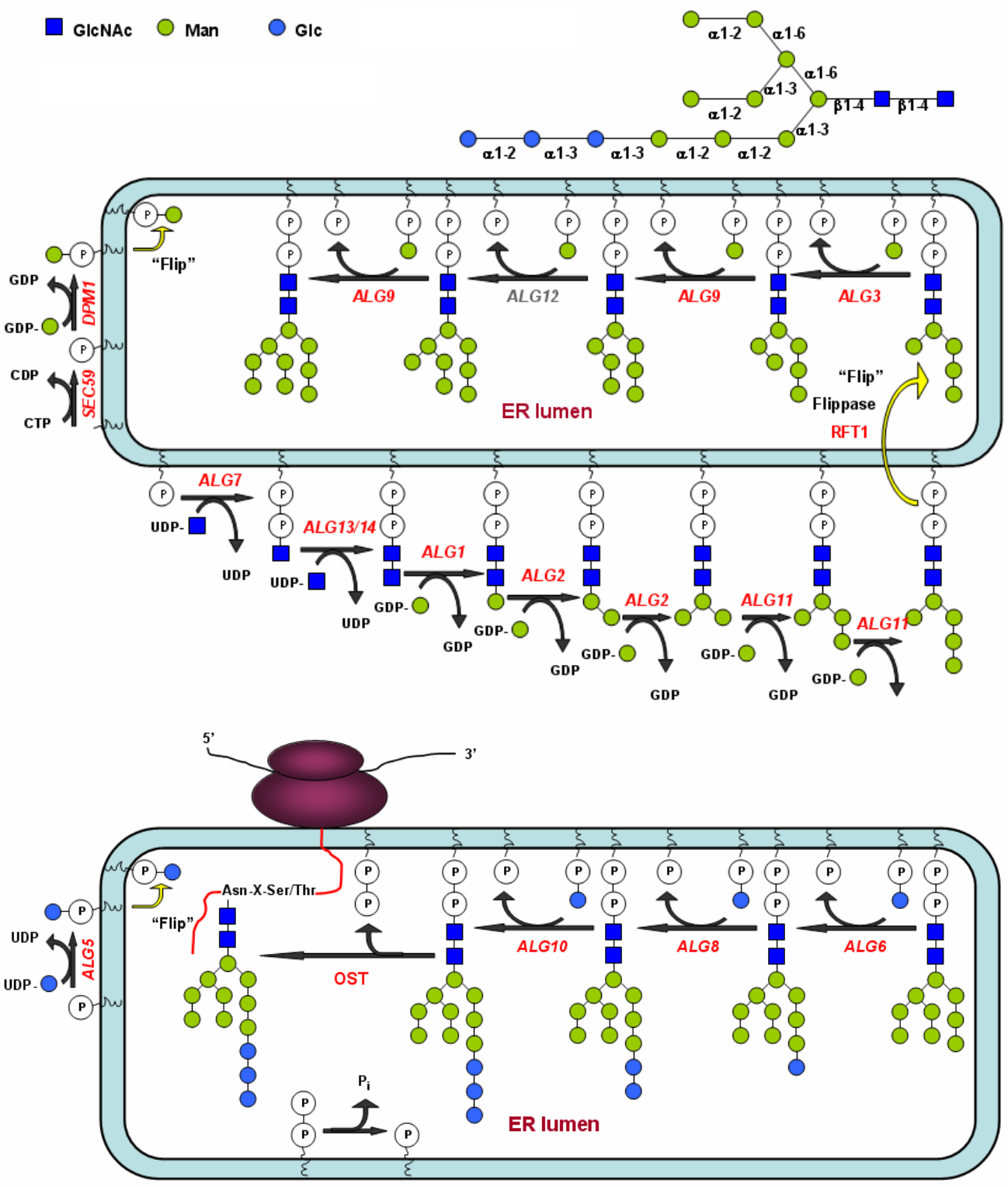


Fig. 1 **Dolichol pathway.** The biosynthesis of the dolichol pyrophosphate-$GlcNAc_2Man_9Glc_3$.

Most O-linked carbohydrate covalent attachments to proteins involve a linkage between the monosaccharide N-acetylgalactosamine and the amino acids serine (Ser) or threonine (Thr). This type of O-glycosylated proteins is also called mucin-type. Currently there is no O-linked amino acid consensus sequence. However, statistical analysis was carried out to study the nature of amino acids around the O-glycosylated Ser/Thr (Thanka Christlet TH *and* Veluraja K, 2001). For this study, 992 sequences containing O-glycosylated Ser/Thr were selected from the O-GLYCBASE database of O-glycosylated proteins (http://www.cbs.dtu.dk/databases/OGLYCBASE/). The frequency of occurrence of amino acid residues around the glycosylated Ser/Thr revealed that there is increased number of proline residues around the O-glycosylation sites in comparison with the nonglycosylated serine and threonine residues. O-linked glycosylation occurs in general at a later stage during protein processing in the Golgi apparatus (with few exceptions, see below). The mucin-type O-glycosylation is initiated by a family of enzymes known as the UDP-*N*-acetylgalactosamine: polypeptide *N*-acetylgalactosaminyltransferases (ppGalNAcTs), which comprises 20 isoforms in humans (only 14 have been functionally characterized to date) (Tian E *and* Ten Hagen KG, 2009). Expression of each family member revealed unique spatial and temporal gene expression patterns during development thus suggesting that acquisition of mucin-type O-glycans is a complex and highly regulated process (Kingsley PD *et al.*, 2000; Tian E *and* Ten Hagen KG, 2006). The role that each ppGalNAc-transferase plays in *O*-glycosylation is unclear (Abeijon C *and* Hirschberg CB, 1987; Wandall HH *et al.*, 1997). Subsequent sugar additions (galactose, N-acetylglucosamine, sialic acid, being the most common sugars) to the core GalNAc residue results in the formation of very complex structures. Such O-glycan structures impart unique structural features to secreted glycoproteins and numerous membrane receptors that are essential components of the extracellular matrix, involved in cell-cell and cell-matrix interactions. Another essential function is to act as a component of mucosal secretions, and it is the high concentration of carbohydrates, which tends to give mucus its "slimy" feel (Varki A *et al.*, 1999). The structural complexity of these chains can sometimes exceed that of N-linked chains by far (Hanisch FG, 2001).

But this is not the only direct form of protein O-glycosylation. Up to now seven other O-linked sugars were found, namely: mannose, fucose, glucose, galactose, xylose, arabinose and GlcNAc.

O-Glycosylation in many fungal species is initiated in the endoplasmic reticulum by protein mannosyltransferases (PMTs), which transfer mannose to serine or threonine residues, and it is completed by mannosyltransferases (Mnt-proteins) in the Golgi (Ernst JF *and* Prill SK 2001). It is

also present in animals and defective O-mannosylations are associated with forms of muscular dystrophy (Lommel M *and* Strahl S, 2009).

O-fucose is added to serine and threonine residues located between the second and third conserved cysteines of EGF-like repeats in the Notch protein, and possibly other substrates by GDP-fucose protein O-fucosyltransferase 1, and to Thrombospondin repeats by GDP-fucose protein O-fucosyltransferase 2 (Moloney DJ *at al.*, 2000; Brückner K *et al.*, 2000). In the case of EGF-like repeats, the O-fucose may be further elongated in mammals to a tetrasaccharide by sequential addition of N-acetylglucosamine (GlcNAc), galactose, and sialic acid, and for Thrombospondin repeats, it may be elongated to a disaccharide with the addition of glucose. Both of these fucosyltransferases have been localised to the endoplasmic reticulum, which was found to be unusual that time, since most of the eukaryotic fucosyltransferases function in the Golgi apparatus (Panin VM *et al.*, 2002; Shao L *and* Haltiwanger RS, 2003; Rampal R *et al.*, 2005). O-glucose is added between the first and second conserved cysteines of EGF-like repeats in the Notch protein by a protein-O-glucosyltransferase (called Rumi) which has been cloned recently (Shao L *et al.*, 2002; Acar M *et al.*, 2008).

O-GlcNAc is added to serines or threonines by O-GlcNAc transferases. In contrast to the above-mentioned O-protein glycosylation forms, O-GlcNAc addition occurs in the cytoplasm and nucleus and is not further elongated. This is a dynamic modification since O-GlcNAc cycles in a manner similar to protein phosphorylation. It appears to occur on serines and threonines that would otherwise be phosphorylated by serine/threonine kinases. In certain proteins, O-GlcNAc and O-phosphate can also occur next to each other (Hart GW *et al.*, 2007). This is an incredibly important finding because phosphorylation/dephosphorylation has become a scientific paradigm for the regulation of signalling within cells. The dynamic interplay between O-GlcNAc and O-phosphate enables the rapid generation of enormous molecular diversity in response to cellular stimuli. So far, more than 500 proteins have been identified to be O-GlcNAc modified (~25% are involved in transcription and translation). O-GlcNAc is for example abundant on proteins involved in stress responses and energy metabolism, or on proteins involved in cytoskeletal regulation. O-GlcNAc is involved also in protein trafficking to the cell surface as well as being a factor slowing protein degradation (Rechsteiner M *and* Rogers SW, 1996; Zachara NE *and* Hart GW, 2006; Hart GW *et al.*, 2007). Furthermore, O-GlcNAc level variations are associated with the development of skeletal muscle atrophy (Konrad RJ *and* Kudlow JE, 2002; Cieniewski-Bernard C *et al.*, 2005). Furthermore,

O-GlcNAcylation of tau protein is much lower in the brain of patients with Alzheimer's disease than it is in a healthy adult brain. O-GlcNAcylation of human tau negatively regulates its O-phosphorylation in a site-specific manner (Liu F *et al.*, 2004; Hart GW *et al.*, 2007).

1.2.2. N-glycosylation in yeasts, invertebrates, plants, and mammals

It has become evident that the N-linked protein glycosylation, which takes place in the ER, and is subject to further processing in the Golgi apparatus, reflect a spectrum of functions related to glycoprotein folding, quality control, sorting, degradation, and secretion. The oligosaccharides not only promote folding directly by stabilizing polypeptide structures but also indirectly by serving as recognition "tags" that allow glycoproteins to interact with a variety of lectins, glycosidases, and glycosyltranferases (Helenius A *and* Aebi M, 2004).

Glycan-protein interactions are of broad diversity as glycan interactions are characterized by graded affinity and multivalency involving multiple protein contacts at various sites of the glycan, thus they are able to regulate pathways in parallel manner and fine-tuning biological responses due to the complexity of their interactions (Collins BE *and* Paulson JC, 2004).

N-linked oligosaccharides in mammals are divided into three different types: high-mannose, hybrid, and complex. While the structural diversity of N-glycans in mature proteins of the cell surface is introduced by glycosyltransferases in the Golgi complex, the core glycans in the ER are universal from yeasts to mammals and have been considered as intermediates (high mannose and hybrid glycans) toward complex glycans. All N-glycans share a common minimal structure *Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc* (Lerouge P *et al.*, 1998).

Yeasts lack hybrid- and complex-type, while producing high-mannose-type N-linked glycans. On the other hand, some high-mannose-type oligosaccharides were described in the case of matured human Tamm-Horsfall glycoprotein (Man_5, Man_6, and Man_7) as well as in the case of human placental arylsulfatase A, even in this case the Man_6 was core α1,6-fucosylated (Serafini-Cessi F *et al.*, 1984; Hoja-Lukowicz D *et al.*, 2000). High-mannose-type oligosaccharides vary in number of mannose units also among yeasts. Even the longest chains produced in *P. pastoris* contain only approximately thirty mannose residues, which is significantly shorter than the 50 to 150 mannose

residue chains typically found on *S. cerevisiae* glycoproteins. The second major significant difference between the glycosylation by *S. cerevisiae* and *P. pastoris* is that glycans from *P. pastoris* do not have α1,3-linked mannose residues that are characteristic of *S. cerevisiae* (Figure 2). The α1,3-mannosyltransferase that makes this linkage, is undetectable in *P. pastoris*. This observation is biologically relevant, because the α1,3-linkages on *S. cerevisiae* glycans are primarily responsible for the highly antigenic nature of glycoproteins used for therapeutic products (Cregg JM *et al.*, 1993).

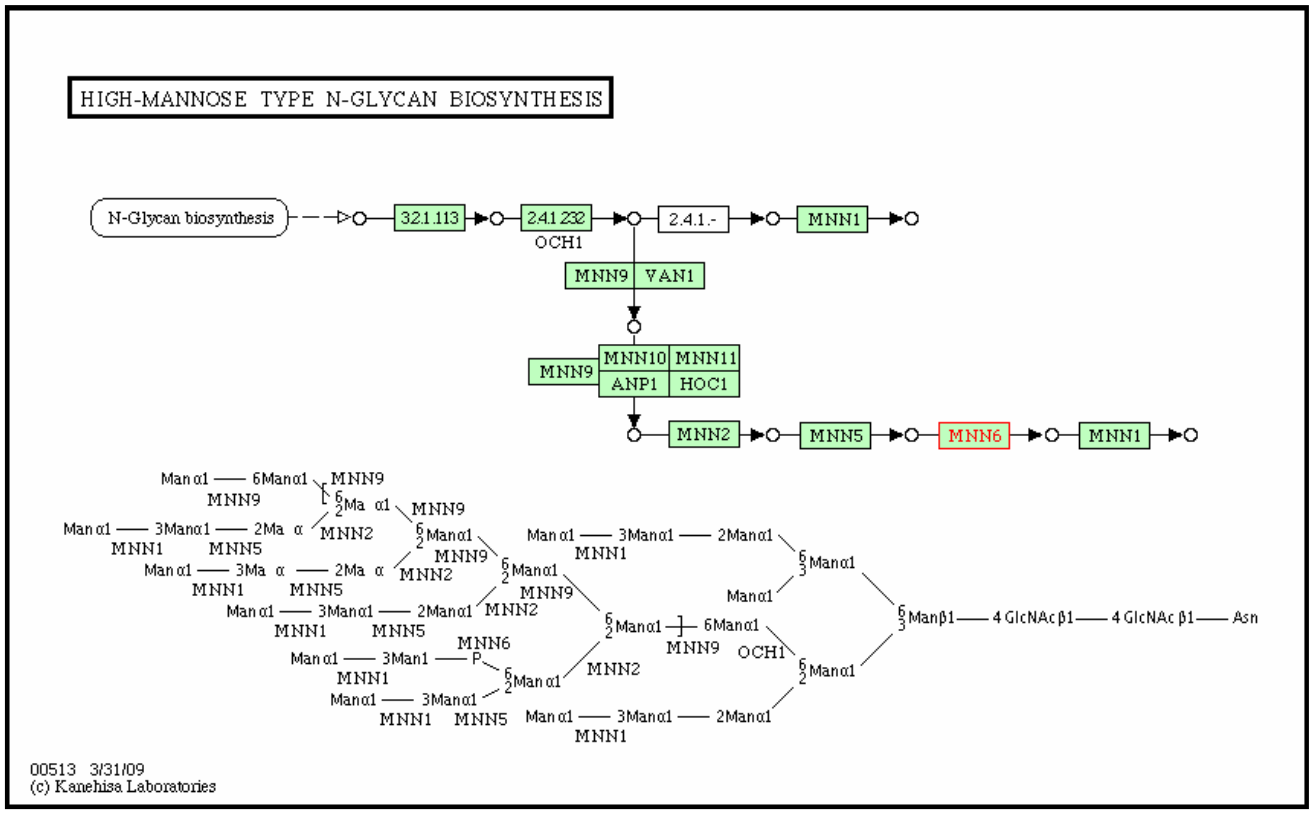


Fig. 2. **High-mannose type N-glycan biosynthesis in *Saccharomyces cerevisiae*.** Genes involved in the biosynthesis are represented in green boxes. Square brackets represent repetitive oligosaccharide sequence. (picture is taken from a webpage of Kanehisa Laboratories http://www.genome.jp/dbget-bin/show_pathway?sce00513+YPL053C)

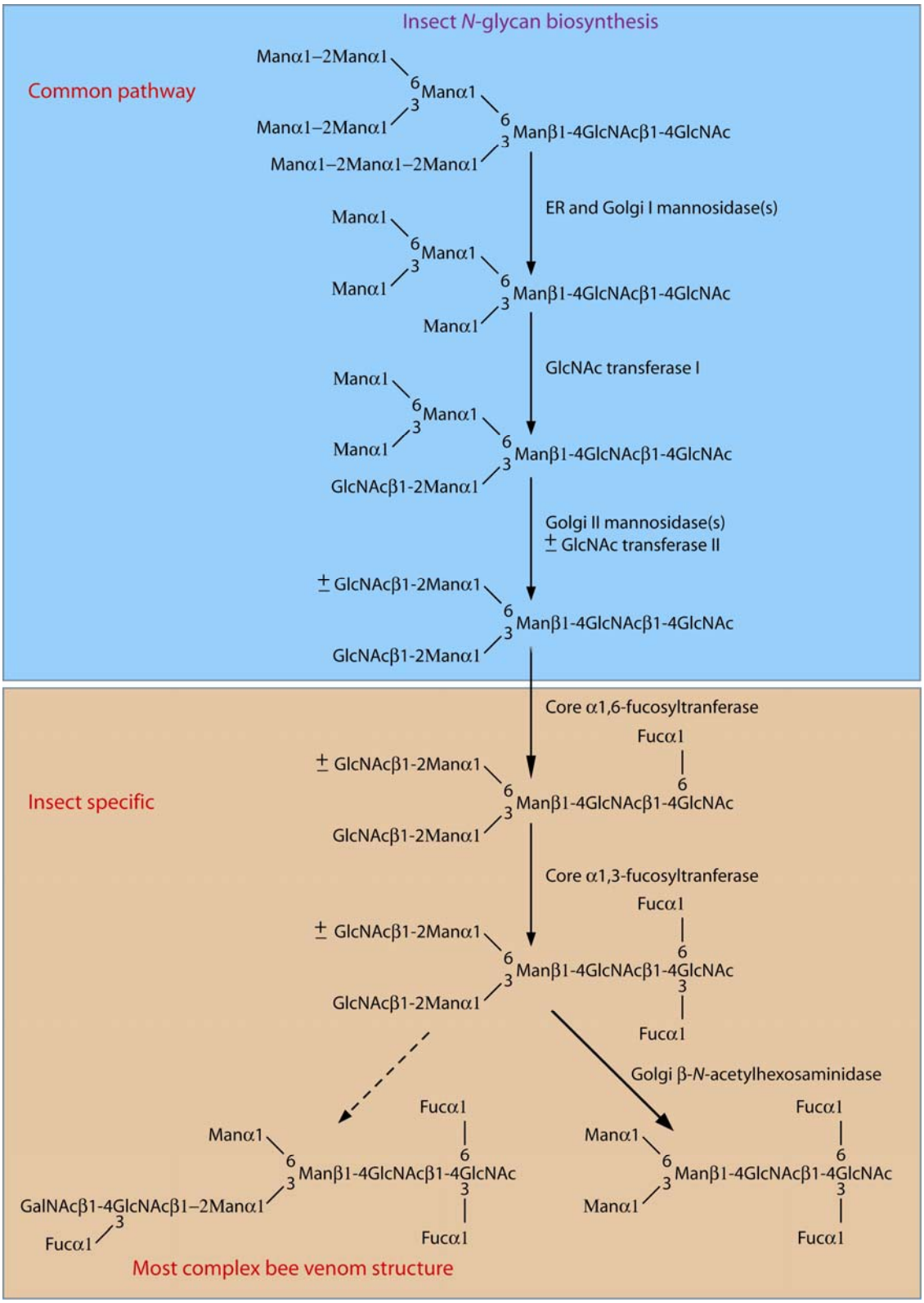


Fig. 3. **N-glycan biosynthesis in insect.** In the blue field one can see the biosynthetic pathway common for mammals, plants, and insects, while the brown field represents the pathway unique to insects. (Wilson IB 2002)

N-glycans found in invertebrates are more similar to those found in plants than to those in mammals. Prof. Wilson and his colleagues did notable research on insect and *C.elegans* glycosylation (Wilson IB, 2002; Paschinger K *et al.*, 2005). They did not find 'complex' N-glycans in invertebrates as they are known in the case of mammals or plants. Only reduced forms were found (the complex character of those N-glycans is given mainly by the presence of fucose residues). In fact N-glycans of *C. elegans* are very similar to those found in insects, but *C. elegans* fucosyltransferases may produce additional structures *in vitro*, even these structures were not found *in vivo*: *GalNAcβ1-4GlcNAcβ1-2Manα1-6(GalNAcβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc* to give *GalNAcβ1-4GlcNAcβ1-2Manα1-6(GalNAcβ1-4(Fucα1-3)GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc,* that was not found in insects. As it is described in figure 3, insect complex N-glycans have reduced biantennary structures with the presence of core α1,6-fucose or both core α1,6- and α1,3-fucose. In the case of the most complex insect N-glycan structures observed in bee venom, a Lewis type α1,3-fucose is present as well as a GalNAc on the non-reducing end.

K. Paschinger and colleagues described four active fucosyltransferases from *C. elegans* (Paschinger K *et al.*, 2004). FUT-1 is a Core type α1,3-fucosyltransferase, which modifies *Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc* on the 'reducing' GlcNAc (Figure 4 and 5). FUT-5 is a Lewis type α1,3-fucosyltransferase, which modifies *GalNAcβ1-4GlcNAcβ1-2Manα1-6(GalNAcβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc* to produce *GalNAcβ1-4GlcNAcβ1-2Manα1-6(GalNAcβ1-4(Fucα1-3)GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc* (Figure 4 and 5). FUT-6 is another Lewis type α1,3-fucosyltransferase, which modifies *Galβ1-4GlcNAcβ1-2Manα1-6(Galβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc* to produce *Galβ1-4GlcNAcβ1-2Manα1-6(Galβ1-4(Fucα1-3)GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc* (Figure 4 and 5). FUT8 is a Core type α1,6-fucosyltransferase, which preferentially modifies *Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4(Fucα1-3)GlcNAc* on the 'reducing' GlcNAc (Figure 4 and 5) (Paschinger K *et al.*, 2005). In the case of *Schistosoma mansoni* worm there is also an additional type of N-glycan *Manα1-6(Manα1-3)(Xylβ1-2)Manβ1-4GlcNAcβ1-4(Fucα1-6)(Fucα1-3)GlcNAc* (Figure 4) (Faveeuw C *et al.*, 2003). Moreover, α1,3-core-fucose and β1,2-xylose, which are immunogenic in mammals and the glycan moieties of this allergen can constitute an IgE-binding determinant (Wilson IB, 2002).

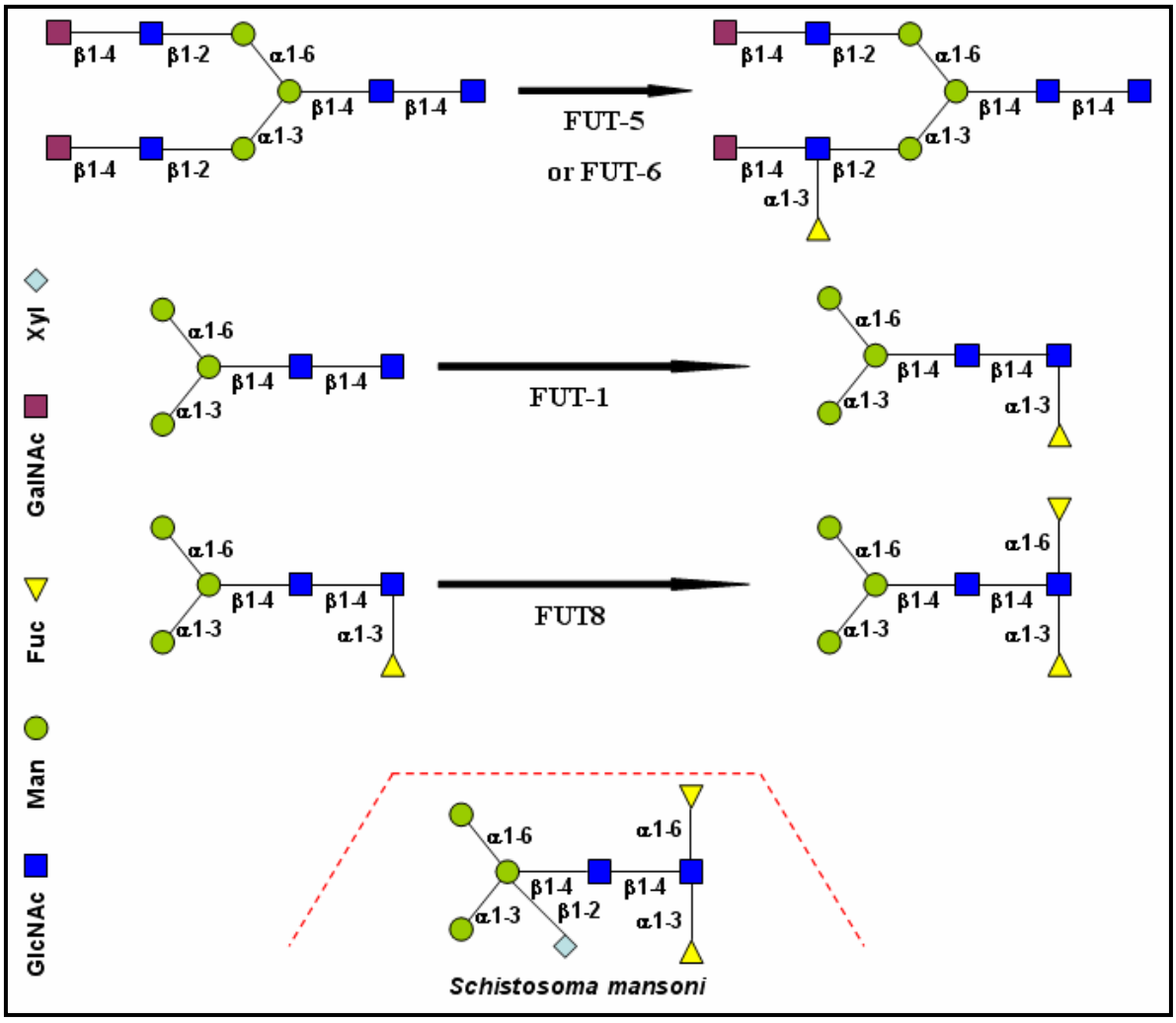


Fig. 4 **Reactions catalysed by *C. elegans* FUTs *in vitro* and a unique glycan structure of *Schistosoma mansoni*.**

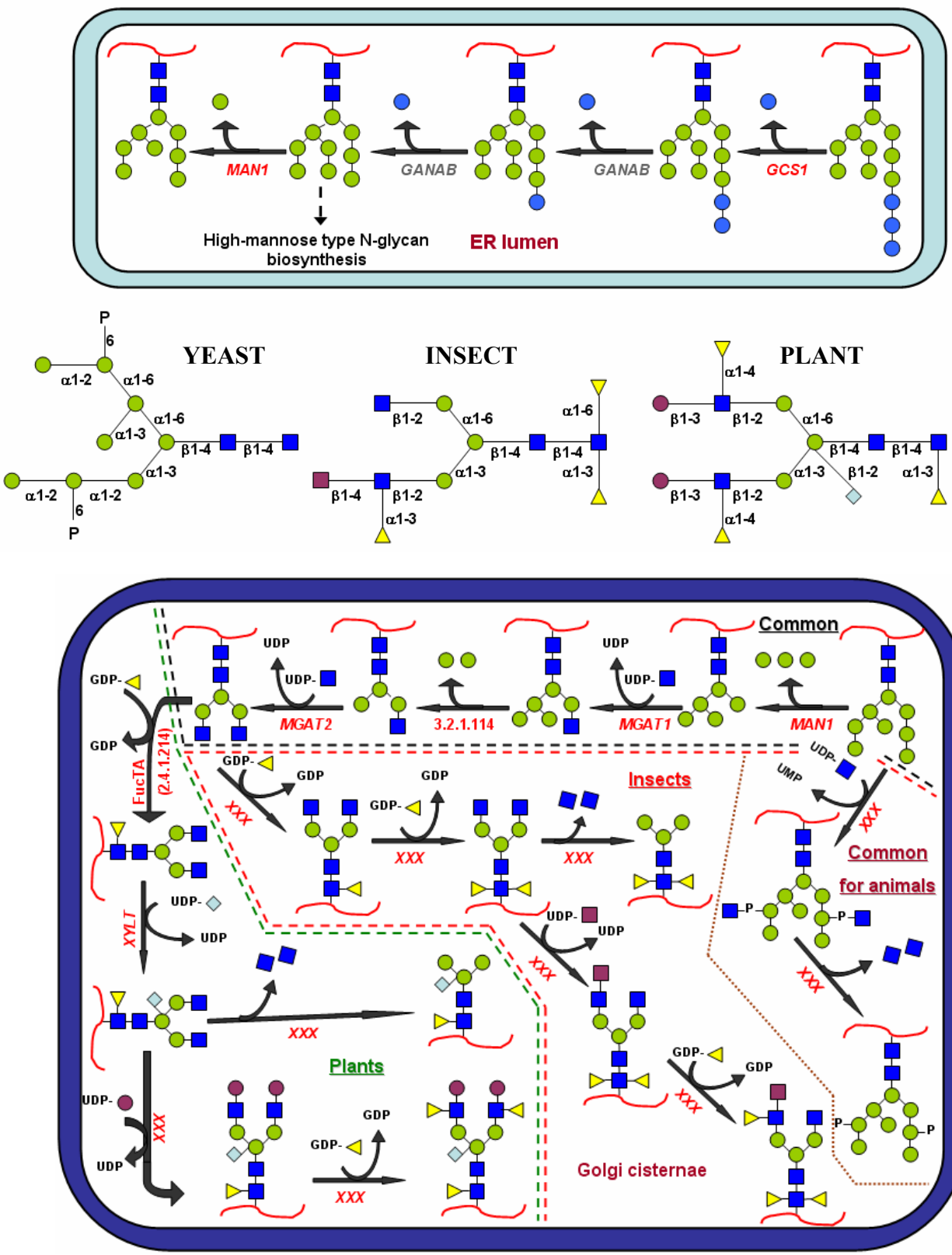


Fig. 5 **Last steps of N-glycan biosynthesis in the Endoplasmic reticulum and N-glycan maturation in the Golgi apparatus.**

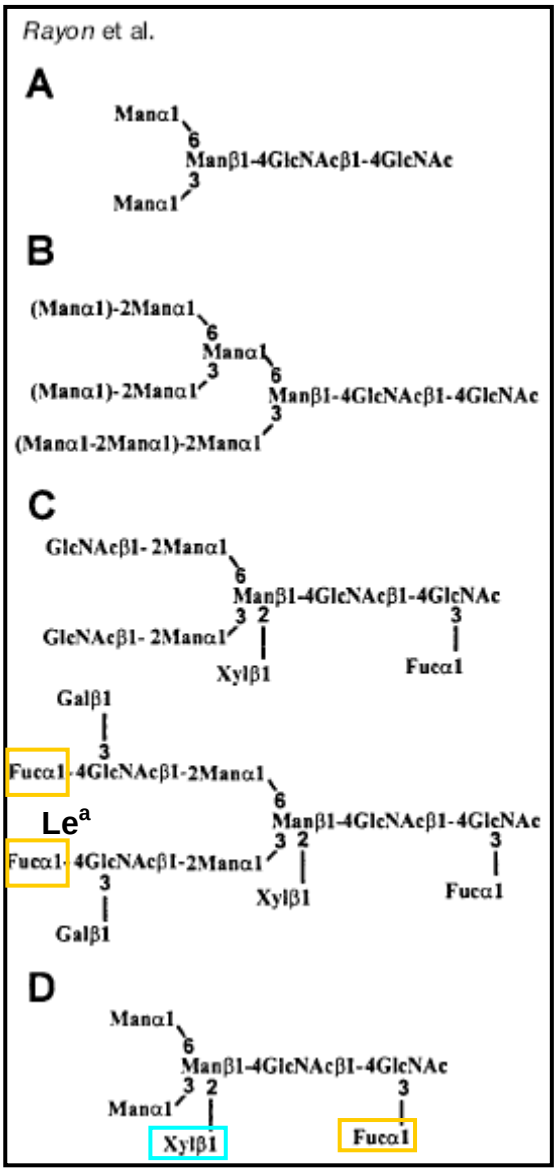


Fig. 6. **Structure of plant N-linked glycans.** (A) Common core of N-linked glycans, (B) high-mannose-type N-glycans, (C) complex-type N-glycans, and (D) hybrid-type N-glycan. (Lerouge P *et al.* 1998)

In plants one can find several types of glycoproteins, some containing oligosaccharides unique to plants, others are similar to those found in fungal and animal systems (Figure 5). Little is known about the addition of sugars to plant-specific glycoproteins such as arabinogalactan proteins (Majewska-Sawka A *and* Nothnagel EA, 2000) or hydroxyproline-rich proteins (Showalter AM, 1993). Therefore we will discuss here only the N-linked glycans that are attached to specific asparagine residues (Asn-X-Ser/Thr where X can be any amino acid except proline or aspartic acid) in many membrane and extracellular proteins. Almost all N-glycans in eukaryotes share a common minimal structure *Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc* (Figure 6A). According to the substitution of this core, plant N-glycans were classified into two categories: high-mannose glycans (with a stoichiometry $Man_5GlcNAc_2$) and complex glycans that have fewer mannose residues and additional residues of other sugars as for example fucose, GlcNAc, galactose and xylose. From results on plant N-linked oligosaccharides obtained in the 90's, Lerouge *et al.* proposed the redefinition of the plant N-glycan classification into the three following classes: high-mannose type, complex type, and hybrid type (also called paucimannosidic type).

The basic biosynthetic pathway of high-mannose and complex N-linked glycans is common (or highly similar) for all eukaryotes (Lerouge P *et al.*, 1998; Keegstra K *and* Raikhel N, 2001). Some complex plant N-glycans are small, containing only two or three mannose residues as well as α1,3-fucose and β1,2-xylose [*GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3)(Xylβ1-2)Manβ1-4GlcNAcβ1-*

4(Fucα1-3)GlcNAc] (Figure 6D), whereas other plant N-linked glycans are larger and have the Lewis a antigen structure, a larger fucose-containing oligosaccharide (Figure 6C).

Since only α1,6 Core type fucosylation occurs in animals (Paschinger K *et al.*, 2005), plant glycoproteins are often immunogenic in mammals, which is caused by their α1,3-core-fucose (predominantly) and β1,2-xylose. This is the so-called horseradish peroxidase epitope, which causes cross-reactivity with the IgE of some allergic patients (Wilson IB, 2002). Since core α1,3-fucose epitope, which is a conserved feature of plants and invertebrates, is not present in mammals, thus the presence of core α1,3-fucose is recognized by mammalian immune systems as a foreign element, which is an epitope for IgE from the serum of many patients with insect, pollen or food allergy and for IgG directed against immunised plant and insect glycoproteins (Wilson IB *et al.*, 1998; Hemmer W *et al.*, 2001). Bardor and colleagues (2003) demonstrated that about 50% of nonallergic blood donors contains in their sera antibodies specific for core xylose, whereas 25% have antibodies against core α1,3-fucose. These antibodies probably result from sensitization to environmental antigens (Bardor M *et al.*, 2003).

Plants are also used as heterologous expression systems for recombinant proteins with therapeutic features. Even if the level of expression is often quite high, biotechnologists have to face the problem of immunogenicity of such proteins, because of their glycoprotein character. As a consequence, strategies to humanise plant N-glycans have been developed. Three main strategies were used to solve the problem. The first is based on the inhibition of the unwanted endogenous Golgi glycosyltransferases (Knocking out of core α1,3-FucT and β-XylT) (Saint-Jore-Dupas C *et al.*, 2007; Sourrouille C *et al.*, 2008). The second is related to the introduction of "new" glycosyltransferases that will allow the synthesis of human-type complex N-glycans (i.e. introduction of β4-GalT gene, and genes necessary for the synthesis and addition of sialic acid) (Bakker H *et al.*, 2001; Sourrouille C *et al.*, 2008). The third approach involves the retention of the recombinant glycoproteins in the endoplasmic reticulum to avoid addition of immunogenic residues (Lerouge P *et al.*, 2000). In the latter case, only high-mannose type N-glycans are produced.

N-linked glycosylation is the most developed in mammals. Some glycotypes are organ specific and depend also on the developmental state of the organism. N-linked glycans in mammals undergo subsequential modifications, which lead to more branched complex glycans as one can see in figure 9. There are also other types of modification as the addition of bisecting GlcNAc (Figure 7 and 10) or Lewis structures (Le^a and Le^x; Figure 8) and Core type fucosylation (Figure 7). In contrast to

plant or invertebrate glycosylation, there is no core α1,3-fucosylation in mammals. On the other hand, β1,4-GlcNAc residues in plants and invertebrates were found neither on mannose residues of the branches nor on the core mannose. Further, there is no evidence of addition of β1,6-GlcNAc to the α1,6 linked mannose in plants and invertebrates (Varki A *et al.*, 1999).

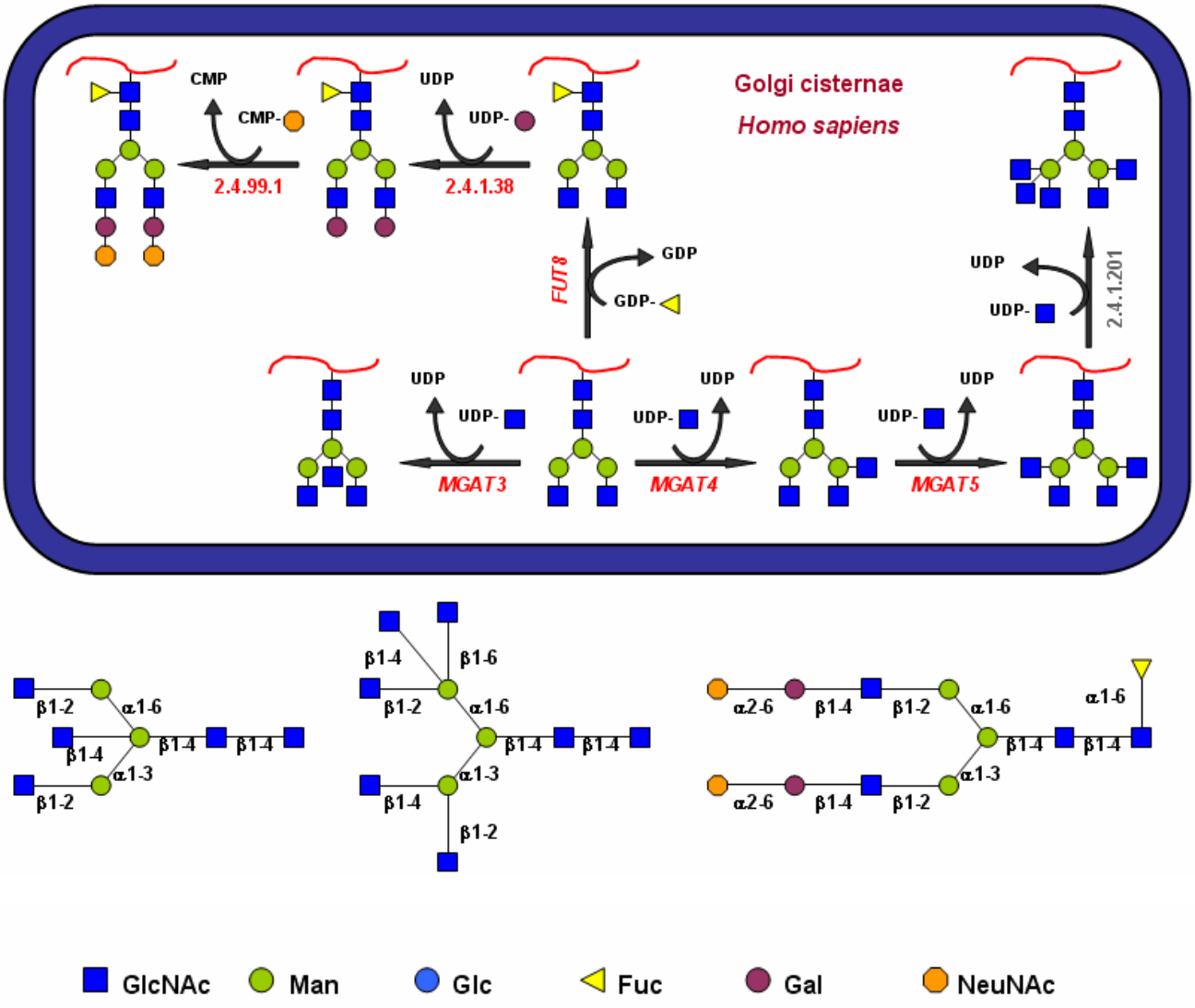


Fig 7: **N-glycan maturation in the Golgi apparatus of *Homo sapiens*.**

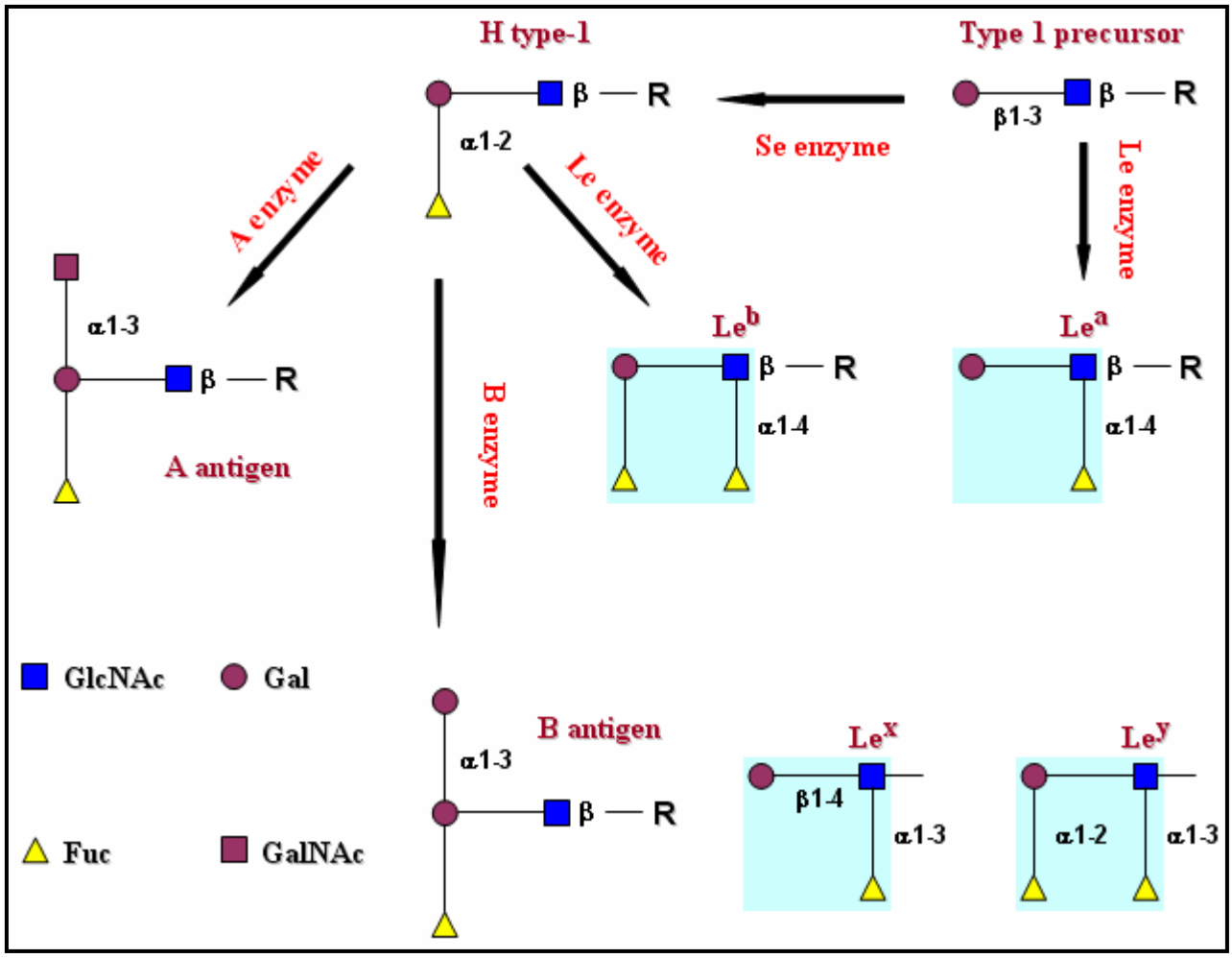


Fig. 8 **Biosynthetic pathways involved in synthesis of Lewis blood group antigens.** Further, Lewis epitopes (Le^a, Le^b, Le^x, and Le^y) are highlighted in turquoise squares.

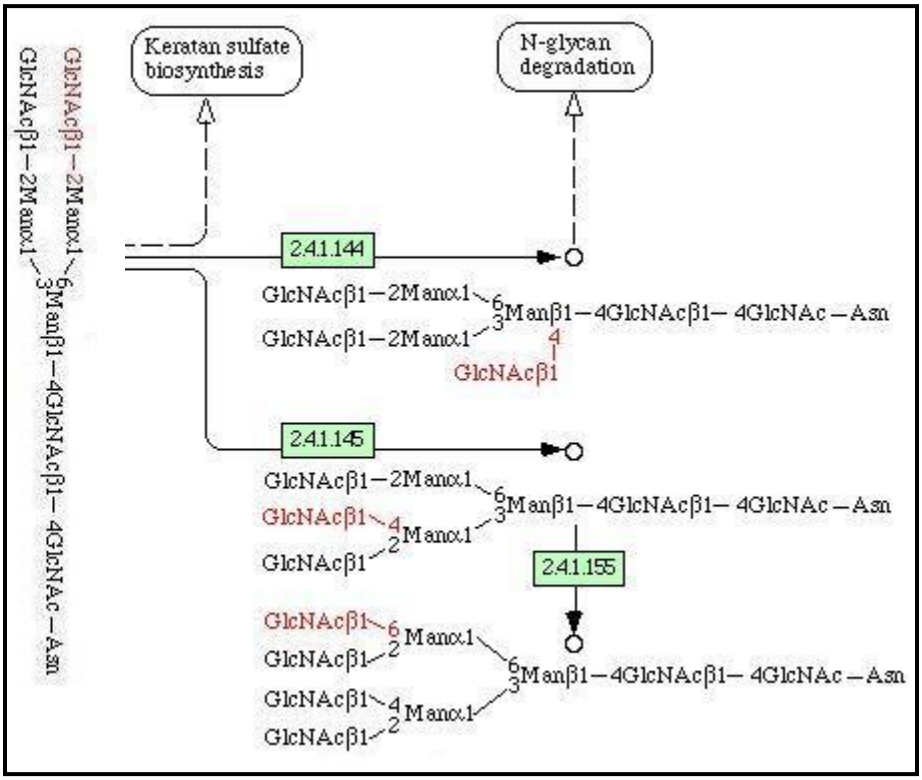


Fig. 9. **Biosynthesis of complex N-glycans in mammals** (the picture is taken from a webpage of Kanehisa Labratories http://www.genome.jp/dbget-bin/get_pathway?org_name=hsa&mapno=00510).

The N-linked glycans listed here are further modified by galactosyltransferases, fucosyltransferases and might be subsequentially sialylated.

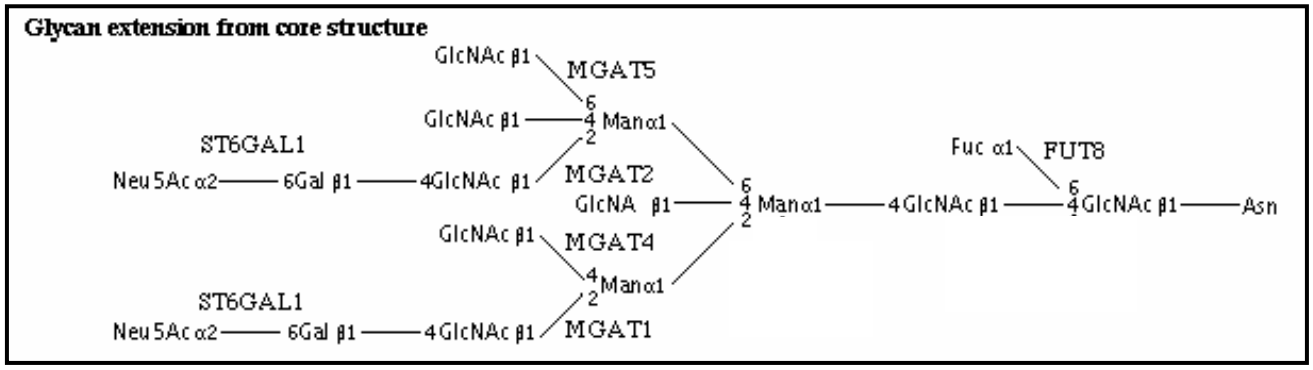


Fig. 10. **Complex N-glycan structure of *Homo sapiens*.** (picture taken from a webpage of Kanehisa Labratories http://www.genome.jp/dbget-bin/get_pathway?org_name=hsa&mapno=00510)

1.3. Role of fucose and disorders associated with altered fucosylation

Fucose plays an important role either as the H antigen of the H/h blood group system or in the Lewis determinants (Flowers HM, 1981). Human FUT1 and FUT2 are α1,2-fucosyltransferases responsible for synthesis of the H blood group antigens (Figure 11) (Kelly RJ *et al*., 1995; Larsen RD *et al*., 1990).

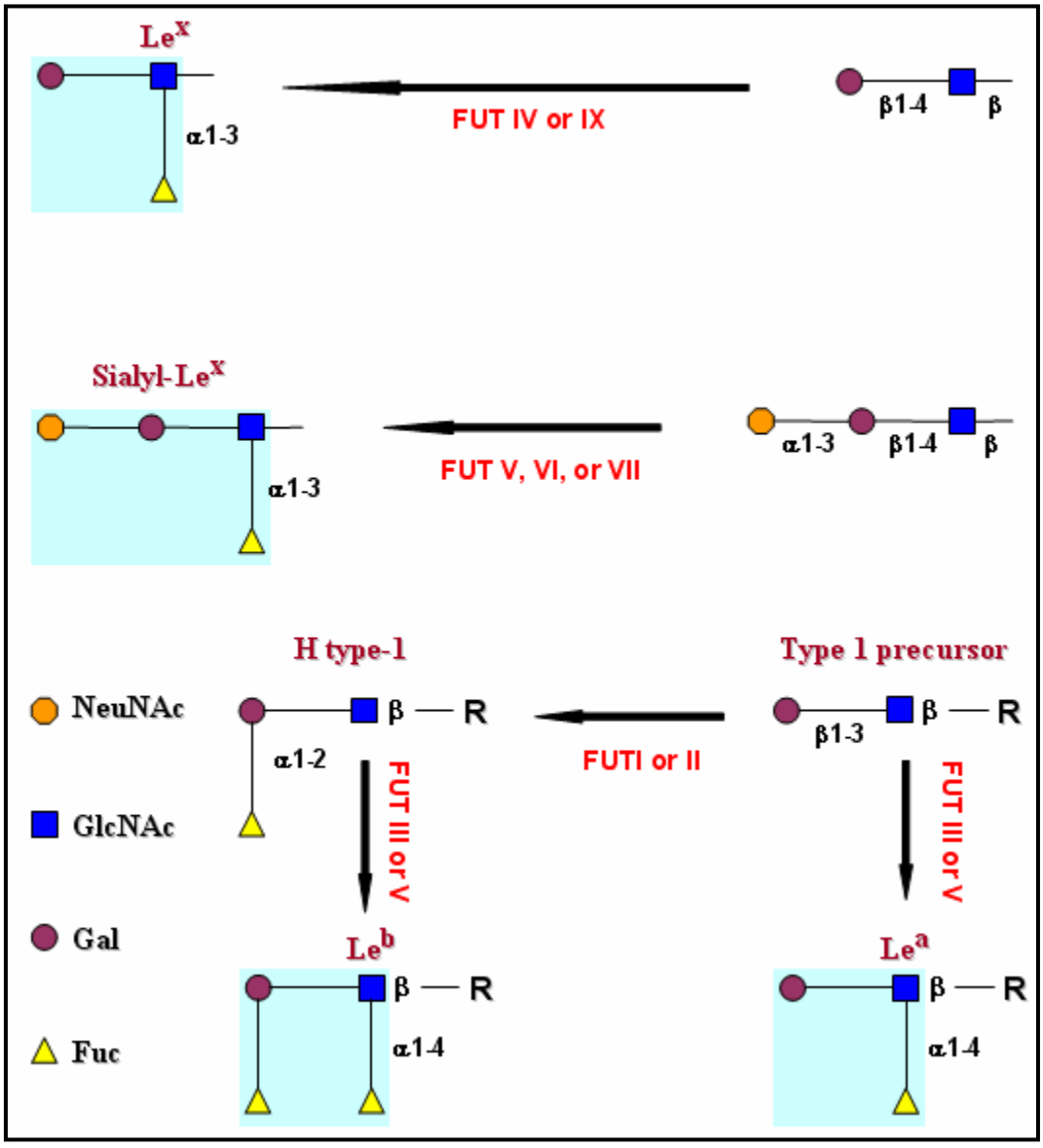


Fig. 11 **Human fucosyltransferases involved in the biosynthesis of the Sialyl-Lewis x (Sialyl-Le^x or sLe^x) and some other Lewis sugars.** Lewis epitopes (Le^a, Le^b, Le^x, and sialyl-Le^x) are highlighted in turquoise squares.

Furthermore, fucose is present in the photoreceptor layer of the retina of the eye where it may be involved in biosynthesis of rod cell glycoproteins (Fliesler SJ *et al.*, 1984). Since fucose is also found in the outer layer of skin, it has been linked to skin hydration functions (Roberts GP *and* Marks R, 1983). Since this sugar is abundant in the proximal tubules of the human kidney, it is most likely important for proper kidney function (Hennigar RA *et al.*, 1985) and because it is found in high concentrations in the testes too, it has been linked to important reproduction functions as well (Malmi R *et al.*, 1987; Domino SE *et al.*, 1989). High concentrations of fucose are also found at the junctions of nerves (Brunngraber EG *et al.*, 1975; Webster JC *and* Klingman JD, 1980) where this sugar plays a fundamental role in nerve cell communication and the extent of protein fucosylation in the brain increases dramatically in response to learning and memory. Inhibition of protein fucosylation using 2-deoxy-D-galactose causes amnesia in animals, presumably by blocking formation of fucose α1,2-galactose linkages. Interestingly, administration of free L-fucose in rats enhances memory retention and long-term potentiation (LTP), a widely accepted cellular model for memory. However, the molecular mechanisms by which fucosyl saccharides stimulate neural connections in the brain are not well understood (Murrey HE *et al.,* 2006). Fucose has also been found to inhibit the ability of bacteria to adhere to cells, without which infection cannot occur (Mason CM *et al.*, 1992). On the other hand, the gastric pathogen *Helicobacter pylori* is capable of attachment to the gastric epithelium via host expression of the Lewis^{b} antigen, a structure containing α1,2- and α1,4-linked fucose that is synthesized by the concerted action of the Se and Lewis (FUT3) fucosyltransferases (Hooper LV *and* Gordon JI, 2001). Furthermore, acidic pH accelerates fucose biosynthesis in *Helicobacter pylori* and since these cells also produce Lewis-related structures, such as Lewis^{x}, Lewis^{y}, and Lewis^{b}, expression of these fucosylated glycans may induce autoimmune-mediated damage to the gastric epithelium, leading to chronic atrophic gastritis in a subset of *H. pylori*-infected humans (Hooper LV *and* Gordon JI, 2001; Moran AP *et al.*, 2002; Becker DJ *and* Lowe JB, 2003).

Patients with rheumatoid arthritis have abnormally fucosylated serum proteins of the acute phase (Flowers HM, 1981; Wiese TJ *et al.*, 1997). Increased expression of fucosylated glycans has also been reported on serum immunoglobins in both juvenile and adult rheumatoid arthritis patients (Flögel M *et al.*, 1998; Gornik I *et al.*, 1999). It is not known if such changes are important to the

pathogenesis of inflammatory arthritis or if they represent a secondary consequence due to upregulation of the fucosylation machinery in the context of autoimmunity.

During the late 70's and 80's numerous studies have demonstrated that many human tumours express fucosylated glycoconjugates that are absent in corresponding normal tissues (Alhadeff JA, 1989; Hakomori S, 1989; Macher BA *et al.*, 1991). Even they did not know what is the functional significance of the expression of these structures in tumours. They inspired others researchers to realise further studies of α1,3-fucosyltransferase activity in cancer patients. In most cases they confirmed increased α1,3-fucosyltransferase activity (Yazawa S *et al.*, 1988 and references therein; Yazawa S *et al.*, 1989). Earlier in 1978, Bauer CH *et al.* showed that a decrease in human serum fucosyltransferase activity is an indicator of successful anti-tumour therapy. Whether the presence of α1,3-fucosyltransferase (it has been shown that fucosyltransferase VI (Fuc-TVI) is responsible for this fucosylation in humans) in serum is a product of tumour growth, or an inflammatory response to tumour growth, was not clear. Recently the FDA has approved AFP-L3, the core fucosylated form of α-fetoprotein as a tumor marker for primary hepatocellular carcinoma (Packer NH *et al.*, 2008). In 2000, Asao *et al.* showed that the elevated expression of Fuc-TVI, has an origin in tumour cells. This enzyme in blood circulation is normally expressed in the liver. Further, Asao *et al.* observed occurrence of the lethal mutations of the FUT7 gene in patients with mental disorders together with presence of new mutated alleles in the Japanese population. Additionally, serum levels of free, non-glycoconjugate fucose are increased in patients with diabetes or cancer (Flowers HM, 1981). On the other hand, fucose inhibited mouse tumor cell-induced platelet aggregation, a process important in cancer cell metastasis (Kijima-Suda I *et al.*, 1988).

In the case of leukocyte adhesion deficiency II fucosylation is gravely reduced, which leads to a compromised immune system (Marquardt T *et al.*, 1999). Interaction between selectins and their ligands enable the rolling of leukocytes on the endothelium, the required first step in leukocyte extravasation (Springer TA, 1994). The carbohydrate selectin ligands are fucosylated structures related to the sialyl Lewis^x structure. Two α1,3-fucosyltransferases, FUT VII (FUT7 gene product) and FUT IV (FUT4 gene product), are expressed in leukocytes and endothelial cells and catalyze the final reaction in selectin ligand biosynthesis, the addition of fucose to sialylated precursors (Lowe JB, 1997; Smith PL *at al.*, 1996)

Involved psoriatic cells retain more of their fucose content within the cytoplasm, whereas normal skin keratinocytes and uninvolved psoriatic cells (the epidermis of uninvolved psoriatic skin

is characterised by a slight hyperproliferation and an increase in inflammatory parameters, whereas no differentiation abnormalities are seen) have most of it on the plasma membrane. This suggests that altered glycoprotein distribution (and metabolism) may play a role in the disease process (Mann PR *et al.*, 1980).

Metabolism of fucose is altered in diabetes. The activities of serum fucosidase and liver fucosyltransferase were increased 100% and 40%, respectively, in diabetic rats (Wiese TJ *et al.*, 1997).

1.4. Fucosyltransferases

1.4.1. Foreword to fucosyltransferases

Many glycoproteins, glycolipids and oligosaccharides contain fucose, which is glycosidically linked to galactose, glucose, N-acetylglucosamine or directly to proteins. Fucosyltransferases are inverting enzymes that transfer fucosyl residues from GDP-β-L-fucose to Gal in an α1,2-linkage, to GlcNAc in α1,3-, α1,4-, or α1,6-linkages (Breton C *et al.*, 1998). Exceptions are the GDP-fucose protein O-fucosyltransferases, which transfer fucose residue directly to the polypeptide chain (Wang Y *et al.*, 2001). The majority of fucosyltransferases need divalent cations for their full activity (Staudacher E, 1996). Since all fucosyltransferases utilise the same nucleotide sugar, their specificity will probably reside in the recognition of the acceptor in relation to the type of linkage formed (Breton C *et al.*, 1998).

1.4.2. Classification of fucosyltransferases

Glycosyltransferases, to which also fucosyltransferases belong, have been classified into different families based on the character of their activated donor substrate (usually a nucleotide-diphospho-sugar), the type of sugar which is transferred, and whether the enzyme forms an α- or β-glycosydic linkage.

In the Carbohydrate Active Enzyme (CAZy) database (http://www.cazy.org/, Coutinho PM *and* Henrissat B, 1999), which is a wide database aimed at classifying enzymes which act on sugars, one can find more than 40000 glycosyltransferase-related sequences. They are divided into 91 families, from GT-1 to GT-91 (plus non-classified sequences) using sequence-based classification (Campbell JA *et al.*, 1997; Coutinho PM *et al.*, 2003; http://www.cazy.org/CAZY/).

On the basis of protein sequence similarities of their catalytic domains, fucosyltransferases mostly classified into 6 GT families (GT10, 11, 23, 37, 65, and 68) according to the CAZy database. However, the presence of three conserved peptide motifs shared by α1,2-fucosyltransferases, α1,6-fucosyltransferases, and protein-O-fucosyltransferases, suggests that they originated from a common ancestor (Breton C *et al.*, 1998; Oriol R *et al.*, 1999; Chazalet V *et al.*, 2001; Martinez-Duncker I *et al.*, 2003). Therefore, two superfamilies have been defined for fucosyltransferases; one containes the α1,3- and α1,4-fucosyltransferases that classify into GT10, and the other includes fucosyltransferase activities that classify into families 11, 23, 37, 65, and 68. All of the fucosyltransferases are expected to share a GT-B fold or a variant of this fold type (Breton C *et al.*, 2006).

The GT-B fold is characterized by two separate Rossmann-type domains with a connecting linker region and a catalytic site located between the domains. Both domains show an α/β/α structure formed by a central parallel β-sheet of topology 321456. This fold is observed in many other GTs, including the prokaryotic enzymes, which produce secondary metabolites (e.g., antibiotics streptomycin, vancomycin) or which contribute to bacterial cell wall biosynthesis (e.g., peptidoglycan) (Breton *et al.*, 2006). The vitally important O-GlcNAc transferase that is responsible for the so-called O-GlcNAcylation of numerous nucleocytoplasmic proteins also belongs to the GT-B fold structural family (Martinez-Fleites C *et al.*, 2008). The C-terminal domain of GT-B enzymes corresponds to the nucleotide binding domain and generally shows an excellent structural conservation. The N-terminal domain is mostly dedicated to the binding of acceptor. Although

peptide motifs characteristics of the GT-B fold, notably a Glu residue and glycine-rich loops, have been previously described (Wrabl JO *and* Grishin NV, 2001), GT-B enzymes do not appear to share any strictly conserved residue (Hu Y *and* Walker S, 2002). Even though divalent cations may be required for full activity of some GT-B enzymes, the mode of activation of fucosyltransferases by certain divalent metal ions is unclear. There is one glycosyltransferase (BGT) adopting a GT-B fold of known complex structure with the presence of divalent metal ion. In this case, divalent metal ions facilitate the release of UDP (Moréra S *et al.*, 2001). For some other glycosyltransferases (a subgroup of the GT1 family) it was shown that the release is facilitated by helix dipole and interaction with some specific residues of the enzyme (Lairson LL *et al.*, 2008).

This is in contrast with another structural family of GTs that adopt a GT-A fold. The GT-A fold consists of an α/β/α sandwich (a mixed seven-stranded β-sheet with 3214657 topology where strand 6 is anti-parallel to the rest) that resembles the Rossmann fold. The first region, encompassing the first 100-120 residues, mostly corresponds to the nucleotide binding domain and is usually terminated by a characteristic Asp-Xxx-Asp (often referred to as DXD motif) (Breton C *et al.*, 1998; Wiggins CA *and* Munro S, 1998). The DXD motif is a degenerate acidic sequence that is shown in all crystal structures to interact primarily with the phosphate groups of the nucleotide donor through the coordination of a divalent cation, typically Mn^{2+}. It has been observed that divalent metal ions, namely Mn^{2+}, play an activator role in the activity of some α3-fucosyltransferases. Although the role of Mn^{2+} is not yet clear, it has been proposed to coordinate the pyrophosphate group in a way similar to that observed in GT-A enzymes (Murray BW *et al.*, 1997; Palma AS *et al.*, 2004). Also alternative divalent metal cofactors were identified as Ca^{2+}, Co^{2+}, and Mg^{2+} (Murray BW *et al.*, 1997).

Recently, three crystal structures of fucosyltransferases belonging to GT10 and GT23 have been solved (Sun HY *et al.*, 2007; Ihara H *et al.*, 2007; Brzezinski K *et al.*, 2007). All of the three structures display a GT-B fold or a variant of this fold.

The fucosyltransferase NodZ (member of the GT23 family) is involved in the biosynthesis of the nodulation factor in nitrogen-fixing symbiotic bacteria (*Bradyrhizobium sp.* WM9). It catalyses α1,6 transfer of L-fucose from GDP-Fuc to the reducing residue of the synthesized Nod oligosaccharide. The enzyme is arranged into two domains of nearly equal size. Although NodZ falls in one broad class (GT-B) with other two-domain glycosyltransferases, the topology of its domains deviates from the canonical Rossmann fold, with particularly high distortions in the N-terminal domain. Mutational data combined with structural and sequence alignments indicate residues of

potential importance in GDP-fucose binding or in the catalytic mechanism. They are all clustered in three conserved sequence motifs (Figure 12) located in the C-terminal domain (Brzezinski K *et al.*, 2007).

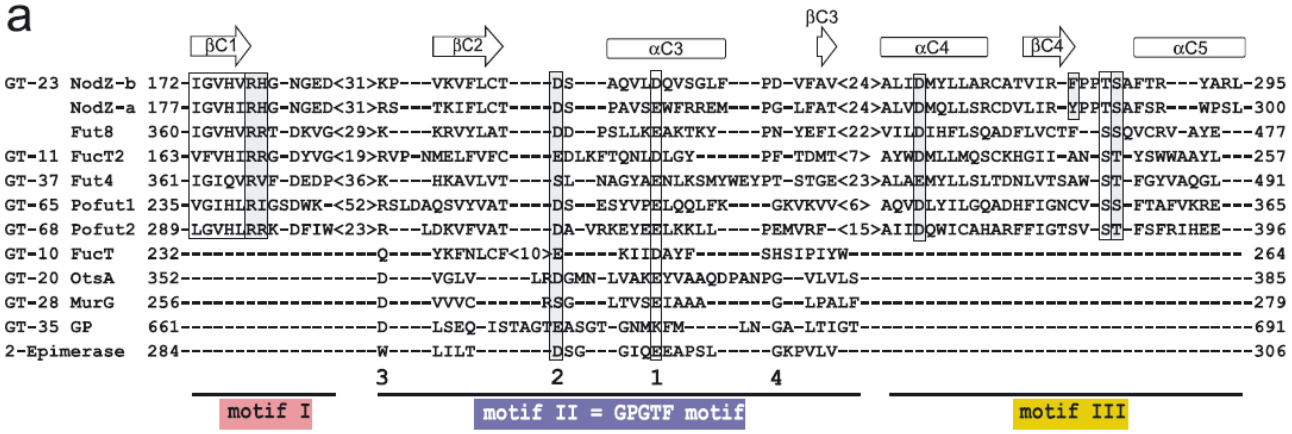


Fig. 12 **Conserved sequence motifs located in the C-terminal domain of NodZ fucosyltransferase (a) and its location in the 3D protein structure (b).** Picture adopted from Brzezinski K *et al.*, 2007.

Human α1,6-fucosyltransferase (FUT8) catalyses the transfer of a fucose residue from a donor substrate, GDP-Fuc to the reducing terminal N-acetylglucosamine (GlcNAc) of the core structure of an asparagine-linked oligosaccharide. The overall structure of FUT8 was found to consist of three domains: an N-terminal coiled-coil domain, a catalytic domain, and a C-terminal SH3 domain (Figure 13). The catalytic region appears to be similar to GT-B glycosyltransferases rather than GT-A. The C-terminal part of the catalytic domain of FUT8 includes a Rossmann fold with three regions that are conserved in α1,6-, α1,2-, and protein O-fucosyltransferases (Ihara H *et al.*, 2007). The SH3 domain of FUT8 is similar to other SH3 domain-containing proteins, although the significance of

this domain remains to be elucidated (Ihara H *et al.*, 2007). According to Ihara and his colleagues the conserved residues in the three conserved regions participate in the Rossmann fold and act as the donor binding site, or in catalysis, thus playing key roles in the fucose-transferring reaction.

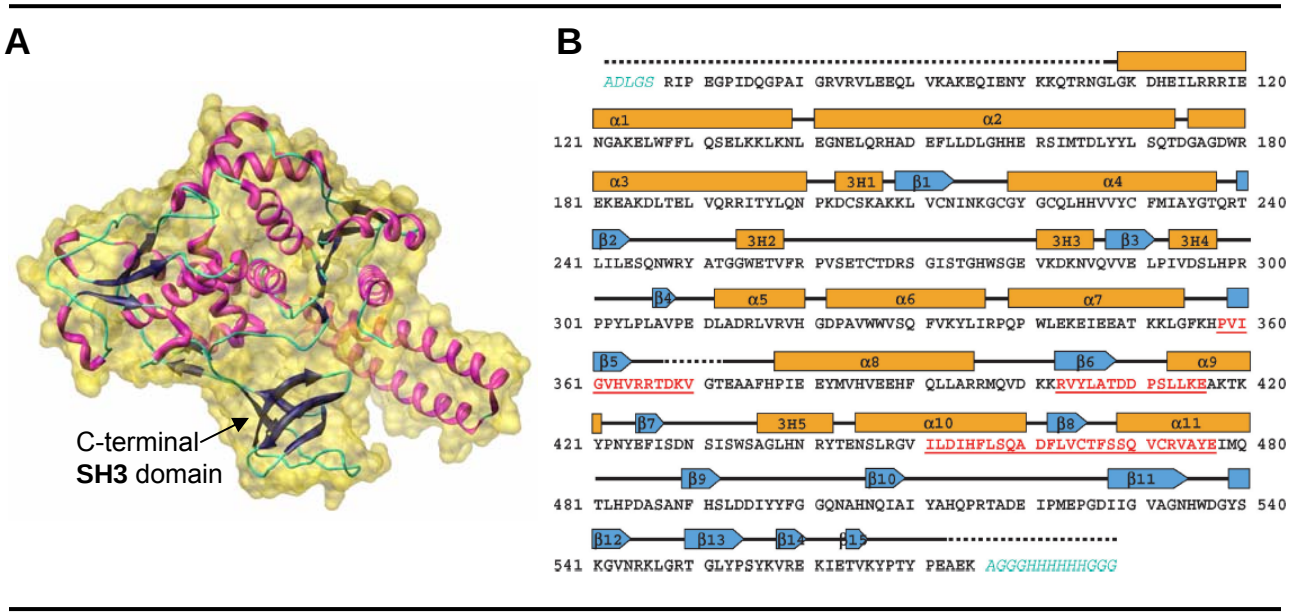


Fig. 13 **3D protein structure of human FUT8 (**ribbon and surface representation; **A) and its amino acid sequence as well as secondary structure (B).** Residues 358–370, 403–416, and 451–477, underlined and in red, indicate the three conserved regions among the α1,2-, α1,6-, and protein O-fucosyltransferases (picture adopted from Ihara H *et al.*, 2007).

Helicobacter pylori α1,3-fucosyltransferase (FucT) is involved in catalysis to produce the Lewis x trisaccharide, the major component of the bacteria's lipopolysaccharides, which has been suggested to mimic the surface sugars in gastric epithelium to escape host immune surveillance. The protein structure is typical of the GT-B family despite little sequence homology (Sun HY *et al.*, 2007). Variations in the protein and ligand conformations, as well as a possible FucT dimer, were also observed. The structure is composed of two similar domains, both having parallel α/β topology of the Rossmann folds (Figure 14). The N- and the C-terminal domains encompass residues 20–150 and 160–320, respectively (Sun HY *et al.*, 2007). The first helix α1(residues 2–13) interacts with the C-terminal domain, whereas the last helix α12 (residues 328–340) interacts with the N-terminal domain.

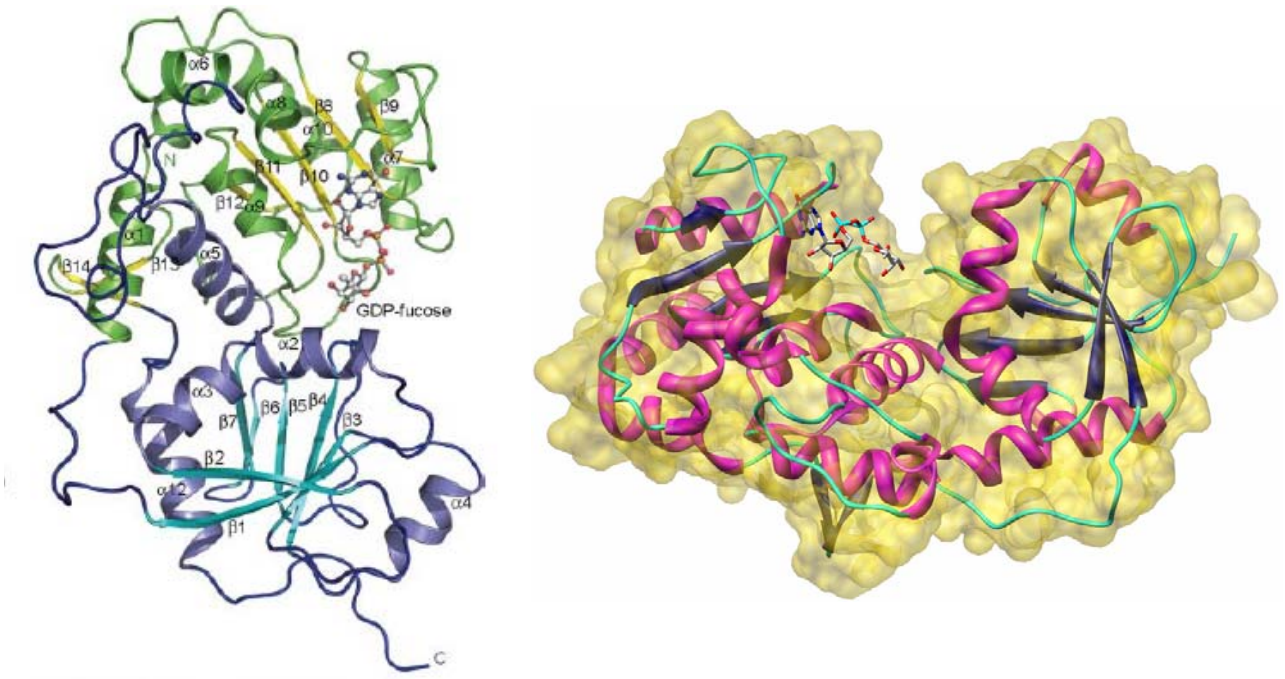


Fig. 14 **Ribbon and protein surface representation of the 3D structure of *Helicobacter pylori* FucT.** On the left side one can see an adopted picture showing numbered secondary structures (Sun HY *et al.*, 2007).

1.4.3. α1,3-fucosyltransferases (Family GT10)

The α1,3-fucosyltransferases transferring fucose to the inner GlcNAc core of N-glycans are called Core type, while α1,3-fucosyltransferases adding fucose to the GlcNAc residues of the branches of N-glycans or other glycoconjugates are called Lewis type (Chen YJ *et al.*, 1998; Wilson IB, 2002). Since α1,3-fucosyltransferases (particularly Core type α1,3-fucosyltransferases) are the subject of the present work, I will further discuss various features of these enzymes.

The eukaryotic FucTs of GT10 family (this is also true for other FucT families) have the typical structure of type II transmembrane proteins (Figure 15), consisting in: (i) a short N-terminal cytoplasmic tail (C), (ii) a transmembrane domain (TM), (iii) a stem region (S), (iv) and a globular C-terminal catalytic domain comprising the conserved peptide motifs (Martin SL *et al.*, 1997;

Grabenhorst E *and* Conradt HS, 1999; Holmes EH *et al.*, 2000; Breton *et al.*, 1998; Chazalet *et al.*, 2001).

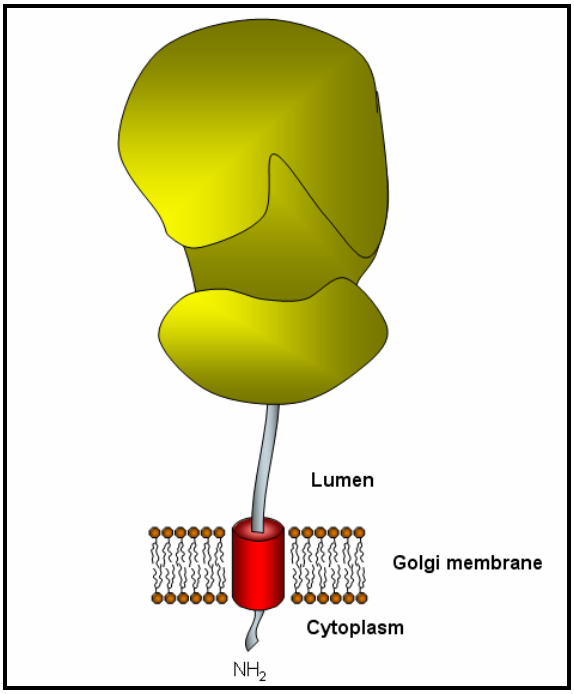


Fig. 15. **General topology of eukaryotic FucTs of GT10 family.** Type II transmembrane protein topology.

Inverting glycosyltransferases, such as fucosyltransferases, transfer a donor sugar to an acceptor substrate by a direct displacement S_N2-like reaction (Lairson LL *et al.*, 2008). S_N2 is substitution nucleophilic bimolecular, according to IUPAC designation it is the type A_ND_N (Figure 16). All fucosyltransferases share the same donor substrate (GDP-fucose).

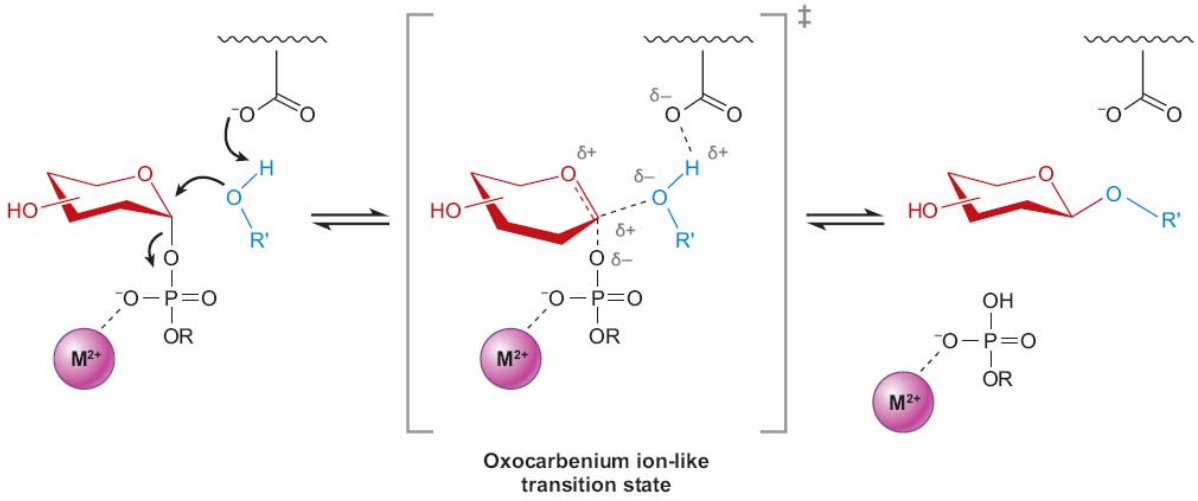


Fig. 16 **Inverting glycosyltransferases utilize a direct-displacement S_N2-like reaction that results in an inverted anomeric configuration via a single oxocarbenium ion-like transition state.** Picture adopted from Lairson LL *et al.*, 2008.

Therefore, it is proposed that all of them will bear conserved amino acids responsible for the donor binding and for the initiation of the reaction, but the identification of these amino acids does not provide sufficient information (e.g., for drug design of specific inhibitors). Therefore solving the structure of even one fucosyltransferase could yield a breakthrough in this field. This was done recently by solving the structure of *Helicobacter pylori* FucT (Sun HY *et al.*, 2007). However, in the case of plant core α1,3-fucosyltransferases primary structure analyses indicate an additional C-terminal subdomain of approximately 110, which does not show similarity to any known protein. The first two plant core α1,3-fucosyltransferases to be characterised either in native or recombinant form (with the contribution of our group) were from *Vigna radiata* (Leiter H *et al.*, 1999) and from *Arabidopsis thaliana* (Wilson IB *et al.*, 2001).

To solve a protein structure, X-ray crystallography is the most suitable method, which requires highly homogenous crystalline protein. Crystallisation is based on the fact that molecules of saturated solution start to crystallize due to slow solvent evaporation. Except of solvent molecules there are other 'helper' molecules in the solvent, which stabilize the protein molecules and have effects on their crystallisation. Co-crystallisation with donor substrate may reveal conformational change in the free and complexed enzyme, but sometimes this is impossible (as well as the

crystallisation step itself). In the last decade, a few rules have emerged to facilitate the structural characterization of eukaryotic glycosyltransferases (Pak JE *and* Rini JM, 2006): expression and crystallization are greatly facilitated by working with a soluble form of the catalytic domain (removal of TM and stem regions); if the protein is heavily glycosylated, removal of carbohydrate chains is recommended since they are usually not uniform and source of microheterogenity. However in the latter case, complete removal of glycan chains may lead to a considerable loss in the solubility of the protein sample. Therefore many crystallographers are looking for the minimal catalytic domain of the enzyme and do screening of glycosylation site mutants or try to express the enzyme in *E. coli* to avoid glycosylation.

The other way to gain insight on the molecular bases of these enzymes are structure-function studies of target enzymes, through protein sequence analysis, fold recognition and homology modelling, site-directed mutagenesis, truncation studies, and comparison of biochemical parameters.

1.4.4. Structure-function studies of α1,3/4-fucosyltransferases

In order to understand structure-function relationships of α1,3-fucosyltransferases as well as to prepare the ground for crystallisation trials, one can refer to previous structure-function studies done on several mammalian FUTs of family GT10. Truncation studies allowed to delineate the minimal catalytic domain of two human α1,3/4-fucosyltransferases, the Fuc-TIII (also called Lewis type enzyme) and Fuc-TV (Xu Z *et al.*, 1996). Sixty-one and 75 amino acids could be eliminated from the N terminus of FucTs III and V, respectively, without a significant loss of enzyme activity. In contrast, dramatic changes of enzymatic activity occurred when deletions of one or more amino acids to the C-terminus were performed. On the other hand, it is very interesting that a C-terminal fusion form of Fuc-TIII with His-tag was successfully expressed in Chinese hamster ovarian cells without losing activity (Augé C *et al.*, 2000).

Truncation of 75 residues from the N-terminus of human Fuc-TIV gave a fully active enzyme (Sherwood AL *and* Holmes EH, 1999). Deletion mutant Δ1-75 of human Fuc-TV was fully active, while Δ1-76 was inactive. Removal of a single residue from the C-terminus decreased enzyme activity by more than 95%. When two residues were removed, the enzyme entirely lost its activity (Xu Z *et al.*, 1996). It is necessary to mention that the very proximal N-terminal regions of the CTS

domain (region upstream of the catalytic domain comprising the cytosolic, transmembrane and stem regions) are important *in vivo,* for enzyme localisation to the Golgi, its orientation and retention (Grabenhorst E *and* Conradt HS, 1999).

As for many Golgi proteins, eukaryotic fucosyltransferases are N-glycosylated. For instance, human α1,3/4-fucosyltransferases III, -V, and -VI contain two conserved C-terminal N-glycosylation sites (Fuc-TIII: N154 and N185; Fuc-TV: N167 and N198; and Fuc-TVI: N153 and N184). Compared to wild types, Fuc-TV and VI single mutations of the first conserved glycosylation site as well as double mutations led to a complete loss of activity, whereas the N198Q and N185Q mutations in Fuc-TV only decreased the activity more than three times. Fuc-TIII glycosylation site mutant N154Q exhibit ~15%, N153Q/N185Q double mutant ~5%, and N185Q ~38% of the activity. Moreover, tunicamycin treatment, which abolishes N-glycosylation, completely abolished Fuc-TIII enzyme activity while castanospermine treatment, which affects processing of N-linked oligosaccharides, diminished Fuc-TIII enzyme activity to ~ 40% of the activity of the native enzyme (Christensen LL *et al.*, 2000 a; Christensen LL *et al.*, 2000 b).

Human Fuc-TIII uses different acceptor substrate as does human Fuc-TV or human Fuc-TVI. A series of swapping experiments of a region, close to the N-terminus of human Fuc-TIII and human FUT V, which differs in these two enzymes, showed that this region is responsible for the acceptor substrate specificity (Xu Z *et al.*, 1996). Only one other study had been published in which the acceptor specificity of human FucTs has been evaluated with respect to the enzyme amino acid sequence (Legault DJ *et al.*, 1995). Legault and colleages used a domain swap approach with full-length forms of human Fuc-TIII and human Fuc-TVI to identify a discrete amino acid sequence, which distinguishes acceptor specificity (*i.e.,* type 1 *versus* type 2). Based on an extensive set of experiments in which various coding sequences of human Fuc-TIII and human Fuc-TVI were swapped, it was concluded that segments encoding amino acid differences between residues at 105–151 of human FucTIII and residues 104–150 of human Fuc-TVI (referred to as subdomains 4 and 5) influence type 1 acceptor specificity. Thus, the substitution of subdomains 4 and 5 of human Fuc-TVI for the corresponding regions in human Fuc-TIII eliminated type 1 acceptor specificity, whereas the complementary construct (*i.e.,* subdomains 4 and 5 of human Fuc-TIII substituted into the human Fuc-TVI coding region) produced an enzyme with both type 1 and type 2 acceptor specificity. However, it should be noted that the latter chimera had very poor activity compared to the wild-type enzymes. Both of the mentioned segments are located closer to the N-terminus of these enzymes.

Alignment of protein sequences of 15 vertebrate α1,3-fucosyltransferases revealed an arginine residue conserved in all compared enzymes employing type 2 substrate exclusively. In the two fucosyltransferases (Fuc-TIII and Fuc-TV), which can also catalyse the formation of an α1,4-linkage, a tryptophan is present at the equivalent position. The single amino acid substitution Trp111 to Arg in Fuc-TIII was sufficient to change the specificity of fucose transfer from H-type 1 to H-type 2 acceptors (Dupuy F *et al.*, 1999).

The α1,3 FucT motif is found in α1,3/4-fucosyltransferases from different species, in different degrees of conservation, and was identified by Martin SL *et al.* (1997). It is a 19 residues region that contains 10 residues, which were found to be conserved in all investigated members of the family (Oriol R *et al.*, 1999). Alanine screening of this motif led to dramatic changes of enzyme activity (see Table I) (Jost F *et al.*, 2005).

In order to determine the minimal catalytic domain of *A. thaliana* core α1,3-fucosyltransferase FucTA, some amino acid truncation from its N-terminus were previously prepared. The first catalytically active construct of FucTA, which was successfully expressed in *P. pastoris* is lacking 66 residues (Δ1-66). Another mutant lacking 88 residues (Δ1-88) was still active, while a mutant lacking 95 residues (Δ1-95) lost its activity (Bencúrová M *et al.*, 2003). The α-factor signal sequence N-terminally fused in frame of these mutants facilitated their secretion into the conditioned media. As the vector (pPICZαC) used by Bencúrová M and her collegues provides a Myc-tag followed by 6xHis sequence located between the MCS and the stop codon, they initially considered of use this C-terminal fusion in the case of Δ1-66 mutant to facilitate its purification. In fact, this C-terminal peptide fusion resulted in lack of protein expression (Bencúrová M *et al.*, 2003).

In the next three tables all α1,3/4-fucosyltransferase point mutants (Table I), truncation mutants (Table II) and glycosylation site mutants (Table III) are listed.

Table I - α1,3/4-fucosyltransferase point mutants

Enzyme	Note	Publication	Point mutation	Activity (in comparison to the non-mutated enzyme)
hFuc-TIII	-	Mollicone R *et al.* 1994	Leu(L)20 → Arg	100 %
			Ile(I)356 → Lys	<10 %
			Thr(T)105 → Ser	0 %
			Gly(G)170 → Ser	0 %
		Staudacher E 1996	Asp(D)336→Ala	0 %
hFuc-TV	Ala screening of the α1,3 FucT motif	Jost F *et al.* 2005	Thr(T)257 → Ala	<1 %
			Glu(E)258 → Ala	<1 %
hFuc-TVI	patients with mental illness	Tanaka S *et al.* 2001	Glu(E)247 → Lys	0 %
			Tyr(Y)315→Stop	0 %
hFuc-TVII	patients *	Bengtson P *et al.* 2001	Arg(R)110→Gln	0 %
hFuc-TIV	Ala screening of the α1,3 FucT motif	Jost F *et al.* 2005	Tyr(Y)240 → Ala	14 %
			Lys(K)241 → Ala	<1 %
			Phe(F)242 → Ala	<1 %
			Leu(L)244 → Ala	7 %
			Phe(F)246 → Ala	7 %
			Glu(E)247 → Ala	<1 %
			Asn(N)248→Ala	<1 %
			Ser(S)249 → Ala	57 %
			Tyr(Y)254 → Ala	0 %
			Thr(T)256 → Ala	<1 %
			Glu(E)257 → Ala	<1 %
			Lys(K)258 → Ala	0 %

*conserved in all previously cloned vertebrate α1,3-fucosyltransferases

Table II - α1,3/4-fucosyltransferase truncation mutants

Enzyme	Prot. Length (aa)	Publication	Truncation	Activity (in comparison to the native enzyme)
hFuc-TII	361	Xu Z *et al.* 1996	Δ1-51	100 %
			Δ1-54	80 %
			Δ1-59	70 %
			Δ1-61	35 %
			Δ1-66	0 %
			Δ361	0 %
hFuc-TV	374	Xu Z *et al.* 1996	Δ1-64	100 %
			Δ1-69	110 %
			Δ1-74	80 %
			Δ1-75	100 %
			Δ1-76	0 %
			Δ1-77	0 %
			Δ373-374	0 %
			Δ374	<5 %
ratFuc-TIV	**	Aucoin JM *et al.* 1998	Δ1-33	100 %
hFuc-TIV	405	Sherwood AL *and* Holmes EH 1999	Δ1-77	100 %
hFuc-TVI	359	Tanaka S *et al.* 2001	Δ315-359	0 %
***A.thaliana* FucTA**	501	Bencúrová M *et al.* 2003 ***	Δ1-66	100 %
			Δ1-88	100 %
			Δ1-95	0 %
***V. radiata* FucTc3**	510	Leiter H *et al.* 1999	Δ1-63	20 %
***C.elegans* FUT-1**	451	Paschinger K *et al.* 2004	Δ1-33	100 %

**Two natural products of the gene (alternative Start codons in the same ORF)

***Addition of Myc-c/His tag to C-terminus of Δ1-66 inactivates the enzyme.

Table III - α1,3/4-fucosyltransferase glycosylation site mutants

Enzyme	Note	Publication	Glycosylation site mutant	Activity (in comparison to the native enzyme)
hFuc-TIII	2 potential glycosylation sites	Christensen *et al.* 2000	N154 → Q	0 %
			N185 → Q	27 %
			Double mutant N154/N185	0 %
hFuc-TV	2 potential glycosylation sites	Christensen *et al.* 2000	N167 → Q	0 %
			N198 → Q	32 %
			Double mutant N167/N198	0 %
***A.thaliana* FucTA**	4 potential glycosylation sites	Bencúrová M *et al.* 2003 ****	N64 → deleted by deletion Δ1-66	100 %

****Truncated protein Δ1-66 is fully active.

The direct outputs from these works are next:

1. α1,3/4-fucosyltransferases are sensitive to their C-terminus.
2. Some amino acids of the highly conserved α1,3-FucT motif are probably involved in the catalytic mechanism of the enzyme
3. Glycosylation might be essential to keep the enzyme active
4. The sequence responsible for the acceptor substrate specificity of eukaryotic α1,3/4-fucosyltransferases is probably located toward the N-terminus from the 3FucT motif
5. The stem region of Golgi α1,3/4-fucosyltransferases is of variable length.

1.4.5. Expression systems used for heterologous expression of α1,3/4-fucosyltransferases

Expression of recombinant human α1,3-fucosyltransferase Fuc-TIX was tested also in *E.coli* (Stacke C *et al*., 2009). Since its long history of laboratory culture and simple manipulation, *E. coli* plays an important role in modern biological engineering and industrial microbiology (Lee SY, 1996). Considered a very versatile host for the production of heterologous proteins (Cornelis P, 2000), researchers can introduce genes into the microbes using plasmids, allowing for the mass production of proteins in industrial fermentation processes. Genetic systems have also been developed which allow the production of recombinant proteins using *E. coli*. Modified *E. coli* have been used in vaccine development, bioremediation, and production of immobilised enzymes (Cornelis P, 2000). *E. coli* cannot, however, be used to produce some of the more large, complex proteins which contain multiple disulfide bonds and, in particular, unpaired thiols, or proteins that also require post-translational modification for activity (Lee SY, 1996). In the recent past new line of *E.coli* was introduced which is able to provide basic N-glycosylation.

Recombinant human fucosyltransferase Fuc-TIX and *Vigna radiata* core α1,3-fucosyltransferase FucTc3 were expressed in insect cells (Stacke C *et al*., 2009; Both P, unpublished data). Insect cells are often used because of their defined N-glycosylation and, being eukaryotic cells, they provide also other post-translational modifications (Jacobs PP *and* Callewaert N, 2009). There are over 500 insect cell lines that have been developed from more than 125 insect species. Insect cells systems are divided into two groups based on their suitability for stable or baculoviral expression strategies (Hitchman RB *et al*., 2009). S2 cell line is the most known system for stable expression (Schneider cells are derived from *Drosophila melanogaster*) (Galesi AL *et al*., 2008). The most known cell lines used as baculoviral expression systems are as follows: Hi5 also called HighFive cells (*Trichoplusia ni*), Sf9 and Sf21 cells (*Spodoptera frugiperda*) (Sander L *and* Harrysson A, 2007).

Recombinant human fucosyltransferases Fuc-TX and Fuc-TXI were expressed in COS-7 cell line (Mollicone R *et al*., 2009). The COS cell line was obtained by immortalizing a CV-1 (Jensen FC *et al*., 1964) cell line derived from kidney cells of the African green monkey with a version of the SV40 genome that can produce large T antigen but has a defect in genomic replication (Gluzman Y, 1981). When an expression construct with an SV40 promoter is introduced into COS cells, the

vector can be replicated substantially by the large T antigen. Two forms of COS cell lines commonly used are COS-1 and COS-7. These cells are a good choise for heterologous expression of human or mammalian proteins (Gluzman Y, 1981).

Pichia pastoris is a species of methylotrophic yeast. It is frequently used as an expression system for the production of proteins (Li P *et al*., 2007). A number of properties make *Pichia* suited for this task: *Pichia* has a high growth rate and is able to grow on a simple, inexpensive medium. *Pichia* can grow in either shake flasks or a fermenter, which makes it suitable for both small and large scale production (Li P *et al*., 2007). The major advantage of *Pichia* over *E. coli* is that *Pichia* is capable of producing disulfide bonds and glycosylations in proteins. This means that in cases where disulfides are necessary, *E. coli* might produce a misfolded protein, which is usually inactive or insoluble. Compared to other expression systems such as S2-cells from *Drosophila melanogaster* or Chinese Hamster Ovary cells, *Pichia* usually gives much better yields. Recombinant *Arabidopsis thaliana* FucTA, *Drosophila melanogaster* core alpha1,3-fucosyltransferase, and *Caenorhabditis elegans* FUT-1 were successfully expressed in this system (Bencúrová M *et al*., 2003; Paschinger K *et al*., 2004).

1.5. Aims of the project

Our first aim was to investigate the protein sequences of fucosyltransferases (putative as well as those with known activity) belonging to the GT10 family by the tools of bioinformatics with an emphasis on plant and nematode core α1,3-fucosyltransferases. As it was mentioned in section 1.4.2., fucosyltransferases of the GT10 family show some sequence similarities at the protein level and they share the same donor substrate. However, one has to consider that their acceptor substrate specificity may differ. Moreover, different α1,3-fucosyltransferases link fucose to different N-acetylglucosamines of the N-glycan acceptor. In other words they are acceptor substrate and site specific, which should be considered from bioinformatics approach. Since plant core α1,3-fucosyltransferases, compared to other GT10 members, possess a unique additional C-terminal region, and also because our group is interested in plant core α1,3-fucosyltransferases (Leiter H *et al.*, 1999; Wilson IB *et al.*, 2001), we decided to further investigate the FucTA from *Arabidopsis thaliana*.

Based on our results of bioinformatics analyses a number of structure-function studies comprising mutant preparation, heterologous expression, and comparative analyses were proposed. We decided to investigate effects of manipulations to the N- and C-terminus of the enzyme, the role of conserved residues of the motif we have found, and the impact of glycosylation on the enzyme activity.

As mentioned above, the crystallographic structure of FucT from *Helicobacter pylori*, a prokaryotic member of the GT10 family, obtained from crystals soaked with the donor substrate was solved recently (Sun HY *et al*., 2007). This fucosyltransferase, compared to *A.thaliana* FucTA, is acting on a very different acceptor substrate. However it shows homology to *A.thaliana* FucTA (as well as to other GT10 members) in a specific region, which corresponds to the donor substrate binding pocket in the 3D structure of *H.pylori* FucT. Thus, our next step was to build a model of this region of *A.thaliana* FucTA using the tools of homology modelling and then using the results from the model for further structure-function studies.

2. Materials and methods

2.1. Bioinformatics

Since we are interested in core α1,3-fucosyltransferases, which belong to the GT10 family comprising all of the known α1,3/4-fucosyltransferases, we decided to analyse this family with special emphasis on plant core α1,3-fucosyltransferases.

To find protein sequence similarities, conserved motifs, and potentially important residues, which may serve as prospective molecular determinants for structure-function studies of the *Arabidopsis thaliana* FucTA, we have analysed the amino acid sequences of forty nine GT10 members (Carbohydrate Active Enzymes database; http://www.cazy.org/; Cantarel BL *et al.*, 2009) using the tools of bioinformatics such as ClustalW (Thompson JD *et al.*, 2002) and Hydrophobic Cluster Analysis (HCA; Gaboriaud C *et al.*, 1987; Callebaut I *et al.*, 1997). Prediction of transmembrane helices was performed by HMMTOP transmembrane topology prediction server (Tusnády GE *and* Simon I, 2001), and Phylogeny by Molecular Evolutionary Genetics Analysis freeware version 4.0 (Tamura K *et al.*, 2007; Kumar S *et al.*, 2008).

The following protein sequences were retrieved from Carbohydrate Active Enzymes database:

Organism	Enzyme	UniProt codes
Helicobacter pylori	FucT	O30511
	FucU	Q9ZKD7
Apis mellifera carnica	FucTA	Q05GU3
	FucTB	Q05GU2
	FucTC	Q05GU1
Arabidopsis thaliana	FucTA	Q9C8W3
Bos taurus	FUTB	Q9TQQ3
	FUT IV	Q8HZR3
Caenorhabditis elegans	FUT-1	Q21362
	FUT-6	Q22202
Canis lupus familiaris	FUT VII	Q659L2
Cricetulus griseus	FUT IV	Q9JIK2
	FUT VIB	Q9R219
	FUT IX	Q9JIG1
Danio rerio	zFT1	Q9W640
Drosophila melanogaster	FucTA	Q9VUL9
Gallus gallus	CFT1	Q98952
	FUT X	Q8AWB5
Glossina morsitans	Core3FT	Q70G57
Gorilla gorilla	FUT V	Q8HYJ7
Homo sapiens	FUT III	Q9P1W6
	FUT IV	P22083
	FUT V	Q11128
	FUT VI	Q9UND8
	FUT VII	Q11130
Homo sapiens	FUT IX	Q9Y231
	FUT X	Q8IVI6
	FUTXI	Q8IYE4
Hordeum vulgare	Core3FT	Q70G73
Hylobates lar	FUT V	Q8HYJ3
Medicago sativa	MsFTa	Q5DTC8
Mus musculus	FUT IV	Q11127
	FUT VII	Q11131
	FUT IX	O88819
	FUT X	Q8C457
	FUT XI	Q8BHC9
Oryza sativa	FUT XI	Q6ZDE5
Physcomitrella patens	FUT	Q6A2M3
Rattus norvegicus	FUT IV	Q62994
	FUT VII	Q712G6
	FUT IX	Q99JB3
	FUT X	Q5F2L1
	FUT XI	Q8CG40
Salmo salar	FUT VII	B5X4F4
	FUT IX	B5X7P5
	FUT XI	B5X1U7
Triticum aestivum	Core3FT	Q70G72
Vigna radiata	FucT (FucTc3)	Q9ST51
Zea mays	FucT	Q0VH31

If several isoforms were available for the same protein then isoform 1 was taken.

In the cases where there was not available data about the minimal catalytic domains or at least the transmembrane domains of the listed protein sequences, such proteins were in the first step analysed by HMMTOP to determine their transmembrane domains. For further analysis only sequences lacking cytoplasmic and transmembrane domain were used.

Sequences of the above mentioned proteins without their transmembrane and most N-terminal regions, or in the case of available data the sequences of their minimal catalytic domains were aligned by ClustalW. Multiple sequence alignment was further processed (fine-tuned) manually and Molecular Evolutionary Genetics Analysis freeware (Tamura K *et al.*, 2007; Kumar S *et al.*, 2008) was used to construct phylogeny by the neighbour-joining method (Figure 19, page 69).

Phylogeny is especially useful when aligning HCA plots (see section 2.1. and 3.1.1.). For example, if one would like to compare HCA plots of two relatively distant members of the GT10 family, then similarities are found easier if these plots are compared also to plots of some other fucosyltransferases, which lie phylogenetically between the two compared sequences.

In searching for common features and differences among the members of the GT10 family, ClustalW (which generates multiple sequence alignments in respect to sequence similarities; using GONNET weight matrix), and Hydrophobic Cluster Analysis (HCA) were employed.

Since HCA is a less known method, we have provided a short introduction of it (the full text was taken from the webpage http://www.lmcp.jussieu.fr/~callebau/hca_method.html).

HCA is a powerful method to search for common structural features among protein sequences displaying low level of sequence similarities (typically less than 20% of sequence identity). It is based on the use of a bidimensional plot, called the HCA plot and whose principles are illustrated in figure 17. The bidimensional plot is associated with an α helicoidal pitch [3.6 residue/turn, connectivity distance (residues separating two different clusters) of 4], which has been shown to offer the best correspondence between hydrophobic clusters (that are circled) and regular secondary structures (Woodcock S *et al.*, 1992; Callebaut I *et al.*, 1997). Examination of the HCA plot of a protein sequence allows to easily identify globular regions from non globular ones and, in globular regions, to identify secondary structures. This 2D signature, which is much more conserved than 1D sequence and which can be enriched from the comparison of families of highly divergent sequences, facilitate the detection, at low levels of sequence identity, of significant similarities (on the structural and functional levels) as compared to background noise.

As one can see in figure 17, the protein linear sequence (1D) (here the human α1 antitrypsin) is shown on the top of the figure with hydrophobic amino acids (V, I, L, M, W, F, Y) coloured. This sequence is written on an α helix displayed along a cylinder. The cylinder is then cut parallel to its axis and unrolled in a bidimensional diagram (2D). This diagram is compacted and duplicated in order to restore the full environment of each amino acid. Hydrophobic amino acids are not distributed randomly but form clusters. The positions of these clusters have been shown to correspond to the positions of regular secondary structures (α-helices and β-strands) (Woodcock S *et al.*, 1992). This is strikingly illustrated by the corresponding experimental structure (3D). The form of the clusters is generally indicative of the type of secondary structures (vertical clusters are often associated to β-strands whereas horizontal ones often correspond to α-helices).

Conversely, sequences stretching between clusters mainly correspond to loops. The 2D structure of a protein sequence can be therefore deduced from the examination of the HCA plot.

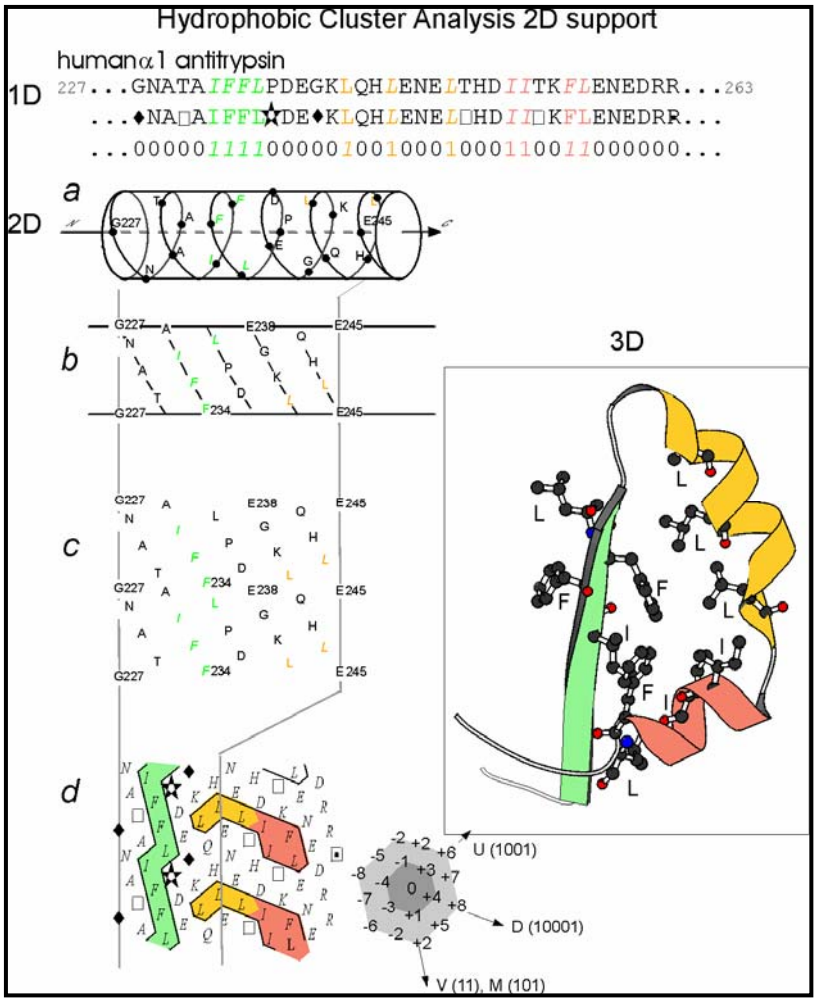


Fig. 17 **Illustration of the principles of the HCA diagram** (Adapted from figure 1 in Callebaut I *et al.*, 1997).

2.2. Media recipes

Percentages are in weight/volume if it is not noted otherwise

LB:
1% tryptone, 0.5% yeast extract, 1% NaCl
Adjusted to pH 7.5 with NaOH.
Sterilised by autoclaving for 20 minutes on liquid cycle.
To prepare plates, 2% agar was added prior to autoclaving.

LB/low salt:
1% tryptone, 0.5% yeast extract, 0.5 % NaCl
Adjusted to pH 7.5 with NaOH.
Sterilised by autoclaving for 20 minutes on liquid cycle.
To prepare plates, 2% agar was added prior to autoclaving.

YPD:
1% yeast extract, 2% peptone, 2% dextrose (D-glucose)
For 1liter of YPD 10 g of yeast extract and 20 g of peptone were dissolved in 900 ml of water (including 20 g of agar if making YPD plates) and autoclaved for 20 minutes on liquid cycle. Solution was cooled to ~55 °C and 100 ml of 20% Dextrose added.

YPDS:
1% yeast extract, 2% peptone, 2% dextrose (D-glucose), 1 M sorbitol
For 1liter of YPDS 10 g of yeast extract, 182.2 g of sorbitol, 20 g of peptone were dissolved in 900 ml of water (including 20 g of agar if making YPD plates) and autoclaved for 20 minutes on liquid cycle. Solution was cooled to ~55°C and 100 ml of 20% Dextrose added.

MYC:
1.25% yeast extract, 2.5% peptone, 1.25% casamino acid

For 800 ml of MYC 10g of yeast extract, 20g of peptone, and 10g of casamino acid (Casein enzymaticaly hydrolysed) were dissolved in water and autoclaved for 20 minutes on liquid cycle.

MGYC:
1% yeast extract, 2% peptone, 1% casamino acid, 1.34% yeast nitrogen base, 1% (v/v) glycerol, 0.00004% biotin.
For 1 liter of MGYC, 100 ml of 13.4% yeast nitrogen base (YNB; Invitrogen), 100 ml of 10% (volume/volume) glycerol (sterilised by filtration), and 2 ml of 0.02% biotin (sterilised by filtration) were added to 800 ml of MYC.

MMYC:
1% yeast extract, 2% peptone, 1% casamino acid, 1.34% yeast nitrogen base, 1% (v/v) methanol, 0.00004% biotin.
For 1 liter of MMYC, 100 ml of 13.4% yeast nitrogen base (YNB; Invitrogen), 100 ml of 10% (volume/volume) methanol (sterilised by filtration), and 2 ml of 0.02% biotin (sterilised by filtration) were added to 800 ml of MYC.

2.3. Cloning of *A.thaliana* FucTA mutants

For the study of core α1,3-fucosyltransferase A (FucTA) from *A. thaliana,* a set of gene constructs derived from a *P. pastoris* shuttle vector (pPICZαC from Invitrogen) was prepared to express the different truncation and point mutants in *P. pastoris*. Due to the sequence upstream of the MCS which encodes the α-mating factor signalsequence, the pPICZαC vector (as well as its derivatives like pPICZαFlagC3 and pPICZαHisFlag; see appendix) allows us to obtain the protein of interest secreted directly into the conditioned media. The pPICZαC (and its derivatives) allows inducible expression of secreted proteins under control of the alcohol oxidase promoter.

Truncation mutants of FucTA were prepared by PCR amplification from the construct pPICZαC/Δ1-66 FucTA (Bencúrová M *et al.*, 2003). PCR was performed using relevant pairs of

primers (Table 1 and 2) and DyNAzyme EXT polymerase (Finnzymes) with its buffer containing 15 mM $MgCl_2$, under the following conditions: one cycle of 95°C for 3 min, 35 cycles of 1 min at 95°C, 45 sec at 47–55°C, 1-2 min at 72°C and with a final extension step at 72°C for 7 min. 0.8% low-melt agarose gel was prepared and electrophoresis of the PCR mixtures performed. PCR fragments of relevant lengths were gel purified (Promega Wizard® SV Gel and PCR Clean-Up System) and incubated for 1 h with *Kpn*I at 37 °C prior to addition of *Eco*RI, after which time the digestion was continued for one further hour. Relevant pPICZαFlagC3 and pPICZαHisFlag plasmids were cut under the same conditions as for the PCR products/fragments and were generally treated with calf intestine alkaline phosphatase (Fermentas) during the final half-hour of restriction digestion. PCR fragments were then ligated into pPICZαFlagC3 or pPICZαHisFlag vector for 3 hours at room temperature using T4 DNA ligase (Fermentas) and a standard ligase buffer.

PCR for truncated mutants

Component	**Volume [µl]**
mQ H_2O	40
10x DyNAzyme buffer with 15 mM $MgCl_2$	5
dNTP mix (2.5 mM each)	1
Forward primer (10 pmol/µl)	1
Reverse primer (10 pmol/µl)	1
Template DNA (10 ng/µl)	1
DyNAzyme EXT (Finnzyme)	1
Total volume	50

Point mutants of FucTA were prepared by inverse PCR amplification using pPICZαFlagC3/Δ1-88 FucTA or pPICZαHisFlag/Δ1-88 FucTA as a template and pairs of 5' phosphorylated mutation primers (listed in Tables 1 and 2). To perform the inverse PCR, Pfu Turbo DNA polymerase (Stratagene) with its buffer and the following conditions were used: one cycle of 95 °C for 2 min, 16 cycles of 45 sec at 95 °C, 1 min at 55 °C, and 13 min 40 sec at 68 °C. iPCR products were gel purified (Promega Wizard® SV Gel and PCR Clean-Up System). Purified iPCR products were then ligated for 3 hours at room temperature using T4 DNA ligase (Fermentas) and standard ligase buffer.

Inverse PCR for point mutants

Component	Volume [µl]
H_2O (from the Promega Wizard® SV kit)	36
10x Pfu Turbo buffer	5
dNTP mix (2.5 mM each)	5
Forward primer (10 pmol/µl)	1
Reverse primer (10 pmol/µl)	1
Template DNA (10 ng/µl)	1
Pfu Turbo DNA polymerase (Stratagene)	1
Total volume	50

Table 1

Name	Primers	Vector or template construct	Note
Δ1-88 C-fusion	Del1/EcoRI, AtFTAasEndP	pPICZαC	without stop codon
FΔ1-88	Del1/EcoRI, 2dA Kpn I	pPICZαFlagC3	stop codon
HFΔ1-88	Del1/EcoRI, 2dA Kpn I	pPICZαHisFlag	stop codon
HFΔΔ1-88;388-501	Del1/EcoRI, 388stopFucTA	pPICZαHisFlag	stop codon
FΔ1-88/S218A	S218Afwd, S218Arev	FΔ1-88	1st Cluster; Donor substrate binding pocket
FΔ1-88/R226A	R266Afwd, R266Arev	FΔ1-88	1st Cluster; Donor substrate binding pocket
HFΔ1-88/N219A	N219Afwd, N219Arev	HFΔ1-88	1st Cluster; Donor substrate binding pocket
HFΔ1-88/Y243A	Y243Afwd, Y243Arev	HFΔ1-88	1st Cluster; Donor substrate binding pocket
HFΔ1-88/S253A	N219Afwd, N219Arev	HFΔ1-88	loop; Donor substrate binding pocket
HFΔ1-88/N337A	N337Afwd, N337Arev	HFΔ1-88	N-glycosylation site
HFΔ1-88/T339A	T339Afwd, T339Arev	HFΔ1-88	N-glycosylation site
HFΔ1-88/T422A	T422Afwd, T422Arev	HFΔ1-88	N-glycosylation site
HFΔ1-88/S483A	S483Afwd, S483Arev	HFΔ1-88	N-glycosylation site

Table 2

List of primers

Name	Sequence (5‘→3‘)
Del1/EcoRI	CCGGAATTCGAGAAATGCCAGGAGTGG
AtFTAasEndP	CGGGGTACCACAAAGACAACTTCGAATTTG
2dA Kpn I	CGGGGTACCTTAGACAAAGACAACTTCG
388stopFucTA	GGGGTACCTCAGGGTCGTCTCTTAAACTCAGG
S218Afwd	GCCAATTGCGCCGCTCGGAATTTCCGCC
S218Arev	AATAAAAGCAGCAGCAAGAGCTTTCTCTG
N219Afwd	CTTGCGCCGCTCGGAATTTCCG
N219Arev	CGGAAATAAAAGCAGCAGCAAGAGC
R226Afwd	CCCTGCAAGCTCTTGAAGCC
R226Arev	CGAAATTCCGAGCGGCGC
Y243Afwd	CTGGTGGTTGTCACCGGAATCGGGATGGG
Y243Arev	CAGAATCAATCTTAACATTCGTCTTCATTAAGGCTTCAAGAGC
S253Afwd	CTGTGGAGAAGGTTGAAGCTCTTAAGCACTACAAATTCAGTC
S253Arev	CCCCATCCCGATTCCGGTGACAACCACCATAAG
N337Afwd	CTCAGACGCTAAGATGGAAACATGAAGGCCCTTCAG
N337Arev	CATAGGCGTCAGGGTTATCCGCAAGATACTTCATTTTCTTTGC
T339Afwd	GCGCTAAGATGGAAACATGAAGGCCCTTCAG
T339Arev	CTGATTATAGGCGTCAGGGTTATCCGCAAG
T422Afwd	GCTCTGGAAGCTCTGGAATCTGCGGTTCTTGC
T422Arev	CAGATTTCCATCCTTCAAGAAGATGGATTCCATGTC
S483Afwd	CTCTCAGTACTCACATACAGAGAAACCCTTGTCCCAAATTCG
S483Arev	CTGAATTTCCTTCGAATTTGAAGTTGTAAAGAGCTTGTCTTTGAGTC
AOXIfwd	CTACTATTGCCAGCATTGCTGC
AOXIrev	GCAAATGGCATTCTGACATCC

2.4. DNA electrophoresis

The success of PCR amplification, DNA fragment insertion (ligation into vector DNA), plasmid preparation, and DNA transformation into competent cells (either *E.coli* DH5α or *P.pastoris* GS115) was investigated by DNA electrophoresis, which is an analytical technique used to separate DNA fragments by size.

30 ml of 1% agarose gel (those from several manufacturers were successfully used, e.g., Sigma, Promega, Lachema) in TAE buffer (preparation of 50x TAE buffer: 242 g of Tris, 57.1 ml of glacial acetic acid, and 18.6 g of EDTA dissolved in mQ H_2O and adjusted to a final volume of 1 litre, for electrophoretic separation of DNA fragments 1x TAE is used), containing 1µl of GoldView™ Nucleic Acid Stain (Beijing SBS Genotech Co.) was prepared and poured into the agarose gel unit of Mini Submarine Model HE33 electrophoresis chamber (Hoefer Scientific Instruments) equipped with sample comb. After the gel has solidified the comb was removed and the chamber was filled with TAE buffer. Samples mixed with 6x Loading dye (Fermentas) as well as MassRuler™ DNA Ladder Mix (Fermentas) were loaded into the sample wells. Electrophoresis was performed at constant power 7 V/cm (approximately 80 V).

For gel extraction of DNA fragments (which is a purification step) the same procedure was performed, the only difference is the use of low melting agarose (from several manufacturers e.g., Sigma, Promega, Serva).

2.5. Preparation of $CaCl_2$-competent *E.coli* DH5α cells

Three milliliters of a 10 ml (LB medium with 10 µg/ml Tetracycline) overnight *E.coli* DH5α culture were transferred into 300 ml fresh LB medium (containing 10 µg/ml Tetracycline), which was shaken for 2-3 hours at 37°C to reach an OD at 600 nm of 0.5. The culture was transferred to sterile centrifuge tubes and left on ice for 10 min. After that the cells were centrifuged (15 min, 4°C,

3000x g) and resuspended in 60 ml of 0.1 M $CaCl_2$ solution, stored on ice again for 30 min, centrifuged (5 min, 4°C, 3000x g) and resuspended in 12 ml of sterile 0.1 M $CaCl_2$ and 2 ml of sterile 87% glycerol. The cell suspension was then aliquoted (100 µl). The aliquots were either directly used or frozen in a mixture of dry ice and ethanol and stored at -70°C.

2.6. Heat shock transformation of competent *E.coli* DH5α cells

An aqueous solution of DNA construct (or ligation product) was mixed carefully with 100 µl $CaCl_2$-competent cells. The mixture was incubated on ice for 30 min, then for 1 min at 42°C (water bath) and again on ice for 2 min. 700 µl LB media were added and the mixture was incubated for 1 hour at 37°C. During the incubation period, the transformation mixture was mixed by inverting the tube several times every 5 minutes, and 100µl of the mixture was spread on a LB/low salt Agar-plate with Zeocin (25 µg/ml; Invitrogen) and plates were left overnight at 37°C. The rest of the mixture was then centrifuged (30 sec, 200x g); most of the medium was discarded, and remaining cells were spread on a LB/low salt Agar-plate with Zeocin and left overnight at 37°C. For positive constructs (as the first template pPICZαC/Δ1-66 FucTA), the mixture was used directly to inoculate an overnight culture for a plasmid prep. Plasmid preparations were prepared using QIAprep® Spin Miniprep Kit (Qiagen).

Positive clones were selected after PCR screening with gene specific primers and by sequencing with vector specific AOXI primers (see Table 2) to confirm the reading frame and to check presence or absence of mutations. The PCR screening was performed on cells (picked colonies), which were resuspended in 16 µl of mQ H_2O, boiled for 5 min, and cooled down on ice for 2 minutes. Then the cell debris was spun down (17000x g, 10 min) and the supernatant was used as a template DNA. PCR conditions were the same as for the truncated mutants.

2.7. Preparation of competent *Pichia pastoris* GS115 cells

Preparation had to be started two days before electroporation. A 5 ml YPD pre-culture was inoculated (by toothpick) with GS115 wild- type on day 1, then incubated shaking at 30°C overnight. On day 2, 250 ml of fresh medium (in an unbaffled 1 liter flask) were inoculated with 100 µl of the preculture, and grown at the same conditions overnight. On day 3, OD_{600} of the culture was measured, and the cells harvested when it reached 1.3 – 1.5. The cells were harvested using a cooled (4°C) Sorvall centrifuge with the SLA rotor and sterilised 250 ml centrifuge bottles. The culture was centrifuged at 1500x g (equals cca. 3500 rpm) for 5 minutes then the supernatant was discarded. The pellet was resuspended in 250 ml of sterile water, centrifuged under the same conditions and the supernatant was discarded again. This procedure was then repeated with 125 ml of water, then with 20 ml of 1 M sorbitol. After the sorbitol supernatant was discarded, the pellet was resuspended in 0.3 - 0.5 ml 1 M sorbitol by gentle pipetting it up and down. The final volume of sorbitol used was low in order to get a very thick yeast suspension. The competent yeast cells have to be stored on ice and used the same day.

2.8. Electroporation of *Pichia pastoris* GS115 cells and heterologous expression of *A.thaliana* FucTA mutants

Five to 10 µg of the expression construct plasmid DNA was linearized by *Pme*I (New England BioLabs) according to the supplier's instructions, then purified by ethanol precipitation, resuspended in 10 µl of HPLC grade H_2O, and used to electroporate *P. pastoris* competent cells (GS115).

Eighty µl of competent cells were mixed with the linearized DNA (5 – 10 µg) in an electroporation cuvette. The mixture was incubated on ice for 5 minutes then the cells were pulsed in a BioRad Micropulser according to the manufacturer's guidelines for yeast (*Saccharomyces cerevisiae*). Immediately after pulsing, 1 ml of 1 M sorbitol was added, the contents were transferred to a sterile 15 ml tube and incubated at 30°C without shaking for 1 – 3 hours. Afterwards, the contents were spread on YPDS agar plates containing zeocin (100 µg/ml), that were then incubated

at 30°C for 2 – 3 days until colonies formed. One or more of them were picked with a sterile toothpick and streaked out on fresh plates. From these plates, colonies were then taken for expression.

Selected colonies of recombinant *P. pastoris* were inoculated into 10 ml of pre-culture glycerol-containing MGYC medium containing 100 μg/ml zeocin. After overnight incubation at generally 30 °C with continuous shaking (200 rpm) in baffled flasks, cells were collected by centrifugation at 1500 × *g*. The cells were washed once in 1.34% (w/v) yeast nitrogen base solution and resuspended in (50 ml culture per 250 ml baffled flask) methanol-containing MMYC medium to a starting OD_{600} = 1.5. Every 24 hours, samples of the cultures cultivated at 16 °C (200 rpm, in baffled 250 ml flasks) were removed and extra methanol added to maintain a concentration of 1% (v/v).

2.9. Yeast genomic DNA isolation

In the case where protein expression screening was not successful (neither core α1,3-fucosyltransferase activity nor the presence of protein in the conditioned media by immuno-staining was proved), genomic DNA of the transformed yeasts was isolated and PCR screened for the presence of the insert by insert specific primers (genomic DNA of wild type GS155 cells served as reference). PCR conditions were essentially the same as for the preparation of truncated mutants except that only 30 cycles were used, plasmid specific AOXI primers (see Table 2) were used, and the starting concentration of the template (instead of 10 ng/μl) was 50 times diluted genomic DNA (see below).

Overnight culture (OD_{600} 4 – 5) prepared in YPDS medium containing 100 μg/ml zeocin of the transformed cells picked from a previously tested colony. One ml of the culture was spun at 5000x g in a 1.5 ml Eppendorf tube and the supernatant was discarded. 0.3g of 425 – 600 micron sterile glass beads, 0.2 ml of lysis buffer [10 mM Tris, 1 mM EDTA, 100 mM NaCl, 1% (w/v) SDS, 2% (v/v) Triton X-100, pH 8], and 0.2 ml of phenol:chloroform (1:1) mix were added to the cell pellet. The tube was spun at top speed for 2 minutes, then 0.2 ml of TE buffer pH 8 (10 mM Tris, 1 mM EDTA) was added and spun again for a few seconds. The mixture was spun for 5 minutes at top speed in an

Eppendorf centrifuge. The aqueous (upper) phase was then transferred to a fresh Eppendorf tube and 2 volumes of 100% ethanol were added to the aqueous phase and mixed thoroughly. The mixture was spun again for 3 minutes at top speed. The supernatant was discarded and the pellet washed with 0.5 ml of ice cold 70% ethanol, then spun for 30 seconds. The supernatant was removed and the pellet air dried and resuspended in 30 µl of TE buffer pH 7 (10 mM Tris, 1 mM EDTA).

2.10. Anti-Flag ELISA

To determine relative specific activities of the mutant enzymes with Flag- or HisFlag-tags, measurement of their protein concentrations by ELISA was employed. To draw a calibration curve, commercial Flag-BaP fusion protein (Sigma) was used (an example of the calibration curve is given in Figure 19 A). This procedure was repeated on each MaxiSorb ELISA plate (Nunc). Samples of the reference mutant HFΔ1-88 were loaded to each plate as well.

Up to 10 µl of *Pichia* supernatant or diluted purified enzyme preparations were pipetted into the wells of MaxiSorb ELISA plates (Nunc), then 40 µl of 50 mM Na_2CO_3 pH 9 were added. The plate was then incubated at 37°C for 2 hours (or overnight at 4°C). The cells were emptied, washed once with TTBS [10 mM Tris, 150 mM NaCl, 0.05% (v/v) Tween 20, pH 8] (200 µl), and then blocked at 37°C with TBS (10 mM Tris, 150 mM NaCl, pH 8) containing 0.5% (w/v) BSA (100 µl) for one hour. The wells were then emptied and anti-Flag monoclonal antibodies raised in mouse (M2) from Sigma, 1:5000 in TTBS containing 0.5% BSA (50 µl) were added. After 1 hour incubation at 37°C, the wells were washed three times, then the second antibody anti-Mouse IgG alkaline phosphatase conjugate from Sigma, 1:5000 in TTBS containing 0.5% BSA (50 µl) was added. After incubation and washing as above, 1mg/ml *p*-nitrophenyl-phosphate (*p*NP-phosphate) in 0.1 M Tris, pH 8.5, was added (150 µl), the plate was incubated at 37°C for 1 hour and measured every 10 min, then the reaction was stopped by addition of 5 M NaOH (50 µl). The plate was then analyzed with a SLT Spectra plate reader, using the machine's PNP program. Calibration was done on Flag-BaP protein (Sigma).

2.11. *A. thaliana* FucTA enzyme assays

In order to verify activity of the secreted forms of *A. thaliana* core α1,3-fucosyltransferase enzymatic assays were performed. In the case of FΔ1-88/S218A and FΔ1-88/R226A mutants, culture supernatants were collected by centrifugation of the conditioned media; for all other forms HisTrap affinity chromatography or Affi-Gel Blue ligand-dye chromatography purified enzyme preparations were dialysed against weak 25 mM Tris buffer pH 7 containing 150 mM NaCl (see section 2.12.). The dansyl-modified form of a $GnGnF^6$ glycopeptide, $GnGnF^6$-NST-dansyl, was used as acceptor substrates. Reactions containing 0.05 mM acceptor, 1 mM GDP-Fuc, 10 mM $MnCl_2$, 50 mM 2-morpholinoethanesulphonic acid (MES), pH 6.5 and 4 µl of enzyme preparation of suitable dilution, to avoid substrate conversion higher than 20% (thus staying in the linear range), were performed in a final volume of 10 µl. Enzymatic assay mixtures were incubated for 1 hour at 30°C and analyzed by reversed-phase HPLC using MZ Analysentechnik Hypersil column (250 x 4.0 mm; ODS C18 5 µm) at isocratic 15 minutes run (flow 1.5 ml/min) with 8.5 % of solvent B (solvent B: 95% acetonitrile in water; Solvent A: 0.05% TFA in water). Excitation λ was set on 315 nm and emission λ on 550 nm.

For measurements of divalent metal ion dependence reaction mixtures with and without 10 mM divalent metal chlorides were prepared. For depletion of residual divalent metal ions, reaction mixtures with 10 mM EDTA were prepared.

To establish the temperature optimum reaction, mixtures with a 50 mM MES of pH 6.5 were used. To establish pH optima, reaction mixtures with 50 mM MES, for the pH range 5.5 - 6.8, and 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), for the pH range 6.8 – 9.5, were prepared. To determine the K_m value for GDP-Fuc, reaction mixtures with 0.06, 0.111, 0.2, 0.333, and 1 mM GDP-Fuc were prepared.

To demonstrate that the enzyme is responsible for the formation of a core α1,3 linkage between the fucose and the acceptor substrate, a different reaction mixture was prepared. GnGn-transferrin was used as acceptor substrates. Reactions containing 2 pmol/µl of acceptor, 1 mM GDP-Fuc, 10 mM $MnCl_2$, 25 mM 2-morpholinoethanesulphonic acid (MES), pH 6.8 and 1 µl of sample containing enzyme were performed in a final volume of 5 µl. Reaction mixtures were analyzed for the presence of $GnGnF^3$-Transferrin in the end-product of the enzymatic reaction. Dot-blotting using BioTrace™ NT Nitrocellulose Transfer Membrane (Pall Corporation) was employed. 1 µl of the

reaction mixture was spotted on the membrane and air-dried. The membrane was then rinsed for 5 minutes in TBS and processed as in the case of western blotting (see section 2.14.) except for the primary antibody, which in this case was 1:4000 anti-HRP raised in rabbit (Sigma), the secondary antibody which was anti-Rabbit conjugated with HRP raised in mouse 1:2000 (Sigma), and the AEC 101 staining kit (Sigma).

2.12. Protein purification

To enrich the recombinant *A. thaliana* core α1,3-fucosyltransferase, two kinds of affinity chromatography were performed according to the tagging present at their N-terminus: (i) a dye-ligand affinity chromatography using Affi-Gel® Blue Gel 100-200 mesh matrix (BIO-RAD), (ii) HisTrap™ HP 1 ml (GE Healthcare) affinity chromatography. Affi-Gel Blue purification was used in the case of mutants bearing only a Flag-tag at their N-terminus, while those possessing a 6x His-tag were purified by HisTrap.

After expression, phenylmethylsulphonyl flouride (final concentration 0.1 mg/ml) was added to the culture medium, which was then centrifuged at 1500 × *g* for 5 min. The resulting supernatant was centrifuged again at 45,000 × *g* for 15 minutes to remove insoluble particles. These steps and conditions were common for both types of chromatography used.

All dialysis steps were performed in cellulose membrane dialysis tubing from Sigma with 12k - 13k MWCO. In the case of Affi-Gel Blue chromatography, all steps were performed at 4 °C. The supernatant was dialyzed against the equilibration buffer (25 mM Tris-HCl, pH 6.8) and loaded on the equilibrated Affi-Gel Blue column (6 ml of matrix, which corresponds to 12 ml of slurry, was loaded on a column with 1.5 cm in diameter) with a flow rate 0.5 ml/min. The column was then washed by 50 ml of 25 mM Tris-HCl, 50 mM NaCl, pH 6.8. Protein elution was performed by 25 mM Tris-HCl, 600 mM NaCl, pH 6.8. Fractions of 0.5 ml were collected and absorbance at 280 nm measured. Fractions containing protein were pooled and dialysed against 25 mM Tris-HCl, 150 mM NaCl, pH 7. The rebuffered protein solution was concentrated by Amicon® Ultra-15 10k MWCO (Millipore) spin column to a final volume of approximately 500 µl.

In the case of HisTrap, all steps were performed at room temperature. The supernatant was dialyzed against the equilibration buffer (25 mM Tris-HCl, 150 mM NaCl, 7 mM Imidazole, pH 8.5)

and loaded on the equilibrated Ni-NTA agarose or HisTrap column (flow rate 1 drop per 2 seconds). The column was then washed by 20 ml of 25 mM Tris-HCl, 300 mM NaCl, 7 mM Imidazole, pH 8.5. Protein elution was performed by 25 mM Tris-HCl, 300 mM NaCl, 500 mM Imidazole, pH 8.5. Fractions of 0.5 ml were collected and absorbance at 280 nm measured in each fraction. Fractions containing protein were pooled and dialysed against 25 mM Tris-HCl, 150 mM NaCl, pH 7. The rebuffered protein solution (still in the dialysis tubing) was concentrated by Spectra/Gel^{TM} Absorbent (SPECTRUM Laboratories) to a final volume of approximately 500 µl. Concentration of total proteins in the samples was determined by Qubit^{TM} fluorometer (Invitrogen).

2.13. PNGase F treatment of *A. thaliana* FucTA (HFΔ1-88)

A portion of the partially purified HFΔ1-88 protein was deglycosylated using N-glycosidase F from Roche, using the manufacturer's standard protocol for complete removal of N-glycans.

2.14. Western blotting

Protein SDS-PAGE was performed in a 10% acrylamide gel based on the procedure of Shapiro and co-workers (Shapiro AL *et al.* 1967). The polyacrylamide gel was then rinsed in blotting solution [50 mM Tris, 40 mM glycine, 0.036% (w/v) SDS, 20% (v/v) methanol] for 5 minutes then placed into the transfer sandwich soaked with blotting solution. Proteins were transferred to the BioTrace™ NT Nitrocellulose Transfer Membrane (Pall Corporation) by semidry blotting (the current used was equal to 2.5 mA per 1 cm^2 of the gel; transfer time 45 minutes).

After the protein transfer the membrane was rinsed in TBS buffer (10 mM Tris, 150 mM NaCl, pH 8) for 5 min. The membrane was then blocked with TBS containing 0.5 % BSA for one hour. The blocking solution was discarded, and anti-Flag monoclonal antibodies raised in mouse (M2) from Sigma, 1:5000 in TTBS [10 mM Tris, 150 mM NaCl, 0.05% (v/v) Tween 20, pH 8] containing 0.5% (w/v) BSA, were added. After 1 hour incubation at 37°C, the membrane was washed three times with TTBS, then the second antibody anti-Mouse IgG alkaline phosphatase conjugate from

Sigma, 1:5000 in TTBS containing 0.5% (w/v) BSA was added. After incubation and washing as above, the membrane was washed once with TBS and immersed into staining solution FAST^{TM} BCIP/NBT (Sigma). The reaction was then stopped with 5% acetic acid.

2.15. Tryptic digest and subsequent MALDI TOF MS analysis of *A.thaliana* FucTA mutants

Mass spectral fingerprinting of SDS-PAGE separated proteins was performed. Band of interest was excised from SDS-PAGE gel. Each gel slice was cut into 1mm cubes and washed with 3 band (gel) volumes of water for 15 minutes; then an equivalent volume of acetonitrile (HPLC grade) was added for a further 15 minutes. A shaking platform was used for all washing steps. The supernatant was removed and 3 band volumes of 0.1 M NH_4HCO_3 (BDH) was added to the gel slice for 15 minutes, then the equivalent volume of acetonitrile was added for 15 minutes. Three more band volumes of acetonitrite were added to dehydrate the gel pieces. The supernatant was removed and the gel pieces were dried in a speed-vac. Proteins were alkylated *in situ*, by adding 2 gel volumes of 10 mM DTT in 0.1 M NH_4HCO_3 (BDH) and incubated at 56 °C for 1 hour. The supernatant was removed. Two gel volumes of 50 mM iodoacetamide in 0.1 M NH_4HCO_3 (BDH) were added and incubated in the dark for 45 minutes at room temperature. The supernatant was removed and gel pieces were washed with 3 band volumes of 0.1 M NH_4HCO_3 (BDH) for 15 minutes, then the equivalent volume of acetonitrile was added for a 15 minutes (this step was repeated once). Three volumes of acetonitrite were added to dehydrate the gel pieces. Supernatant was removed and the gel pieces were dried in a speed-vac. Two gel volumes of 12.5 μg/ml modified trypsin (sequencing grade Boehringer) in 20 mM NH_4HCO_3 (BDH) was added to the dried gel pieces and incubated on a shaking platform at 30 °C for 30 minutes, then a sufficient volume of 20 mM NH_4HCO_3 (BDH) was added to cover the gel pieces and incubated at 30 °C for 18 hours. The supernatant was transferred into a fresh tube. An equivalent volume of 50% acetonitrile/0.1% TFA was added and incubated on a shaking platform for a further 30 min. This supernatant was removed and combined with the first

extract, then dried in a speed-vac. Protein sample was reconstituted in 10 µl of 50% acetonitrile, 0.1% TFA, spun, and sonicated in a sonication bath for 5 minutes.

The MALDI TOF MS was performed in Vienna by Ao.Prof. Iain B. Wilson.

3. Results and Discussion

3.1. Bioinformatics

3.1.1. Sequence analysis of α1,3/4-fucosyltransferases

Since we are interested in core α1,3-fucosyltransferases, which belong to the GT10 family comprising of all α1,3/4-fucosyltransferases, we decided to analyse this family by the tools of bioinformatics (see Materials and Methods) and with an emphasis on plant core α1,3-fucosyltransferases.

During the last two decades several studies dealt with bioinformatics analysis of protein sequences of enzymes belonging to the GT10 family. The story started with the discovery of the so-called α1,3-FucT motif (Martin SL *et al.*, 1997) which was also described by Breton and her collegues as motif II beside another motif called I (Figure 18) (Breton C *at al.*, 1998). These motifs were observed during the analysis of the sequences of Fuc-TIII-VII and Fuc-TIX by Hydrophobic Cluster Analysis (HCA). Later a third conserved peptide motif (I/V/F)HH(R/W)(D/E)(I/V/L) was revealed and called the acceptor motif (Figure 18), because replacement of a single residue Trp111 to Arg within this motif of Fuc-TIII led to the formation of Le^x epitope (α1,3-fucosylation, which requires β1,4 bound galactose to the GlcNAc) instead of Le^a epitope (α1,4-fucosylation, which requires β1,3 bound galactose to the GlcNAc) (Dupuy F *et al.*, 1999; Dupuy F *et al.*, 2002). The forth conserved motif, located toward the N-terminus, prior to the acceptor motif, was later called motif III (Dupuy F *et al.*, 2002; Dupuy F *et al.*, 2004). Dupuy and his collegues claim that motif III (Figure 18) is expected to correspond to the beginning of catalytic domain of vertebrate α1,3/4-fucosyltransferases (Dupuy F *et al.*, 2002). The order of these motifs from the N- toward the C-terminus is as follows: III, acceptor motif, I, and II. Recently, a fifth conserved motif which lies between the acceptor motif and motif I was found (Mollicone R *et al.*, 2009). Mollicone and her collegues, in order to avoid confusion, renamed these five motifs from I to V, according their order in the protein sequence.

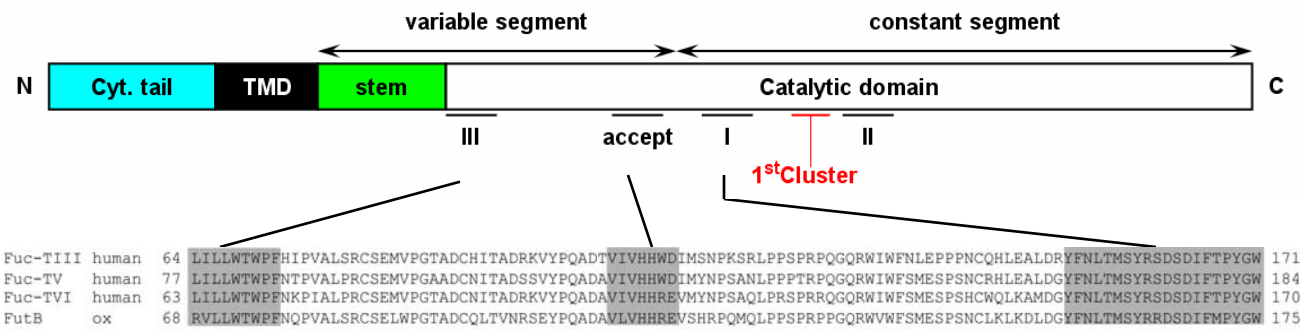


Fig. 18 **Schematic representation of vertebrate α1,3- and α1,3/4-FUTs.** These enzymes are type II Golgi-anchored glycoproteins with a short N-terminal cytoplasmic tail (cyt. tail), a transmembrane domain (TMD) followed by a luminal C-terminal segment, composed of a stem and a catalytic domain. In these enzymes, most amino acid differences are localized in the amino terminal part of these enzymes (variable segment), and their catalytic domains are highly conserved (constant segment). The peptide motifs characteristic of α1,3- and α1,3/4-FucTs are indicated: the motifs I and II are marked domains (Oriol R *et al.*, 1999); motif III could correspond to the beginning of the catalytic domain (Dupuy F *et al.*, 2002); amino acids of the acceptor-binding motif (accept) are involved in the acceptor substrate recognition and specificity (Dupuy F *et al.*, 1999); the 1stCluster motif (subject of this work) and the motif II (α1,3-FucT motif) are involved in the donor substrate binding (Jost F *et al.*, 2005). The alignment of peptide sequences between motifs III and I of enzymes used in this study is revealed. The amino acids involved in type 1 acceptor substrate specificity are indicated by stars (Legault DJ *et al.*, 1995) and arrows (Nguyen AT *et al.*, 1998) (picture partially adopted from Dupuy F *et al.*, 2004).

First we selected sequences of all GT10 members with known function and complemented them with sequences of putative α1,3/4-fucosyltransferases. A phylogenic tree of these sequences was built up to group the chosen sequences and make their comparison easier. As one can see plant core α1,3-fucosyltransferases are related to *H. pylori* FucT and FucU and to Fuc-TXs and Fuc-TXIs of higher eukaryotes (Figure 19). The first conserved sequence **(I/V)(L/M)(L/V)WXXP** among mammalian FUTs (Figure 21) we have found is the so called motif III (or I according to Mollicone; which is not present in the sequences of plant and invertebrate α1,3/4-fucosyltransferases). The first conserved motif among seeding plant core α1,3-fucosyltransferases is **C(E/Q)(E/Q)(W/R)LE** (Figure 21), the corresponding sequence of a nonseeding moss (*Physcomitrella patens*) FucT is **CSAEWLE**. This conserved sequence starts with C91 in *A. thaliana* FucTA. Since the N-terminal truncation of 88 residues did not significantly reduce the enzyme activity, while truncation of 95 residues led to complete loss of activity (Bencúrová M *et al.*, 2003), it is very probable that the

minimal catalytic domain of *A. thaliana* FucTA starts at or just before C91. The so called acceptor binding motif (or motif II according to Mollicone) is present also among the sequences of *C. elegans* α1,3-fucosyltransferases except for FUT-1 which is a core α1,3-fucosyltransferase and recognises a simple acceptor substrate as other core α1,3-fucosyltransferases. Motif I is present in the sequences of mammalian and invertebrate α1,3/4-fucosyltransferases, but not in plant core α1,3-fucosyltransferases. We also confirmed the presence of the α1,3 FucT motif (Martin SL *et al.*, 1997) in all 49 sequences selected for bioinformatic analysis. While proceeding to the extensive protein sequence analysis, we identified a new conserved motif that was tentatively named "1st Cluster motif" (Figures 18 and 20) as it was identified using the HCA method (see chapter 2.1.). This motif is located upstream the α1,3-FucT motif and it can be written as follows: **S-(N/H/D)-X_{5-9}-R-X_6-(L/I)-X_{3-7}-(V/L/I)-X_3-G**. The distance between the two motifs approximately ranges from 17 to 30 amino acids (Figure 20). This motif is more conserved at the level of hydrophobic clusters than at the level of primary structure, which suggests rather a structural conservation. The "1st Cluster motif" consists of three conserved residues and some residues, which have side chains of similar properties. The short region is spanning residues from 218 to 244 of the *Arabidopsis thaliana* FucTA protein sequence (Figures 20, 21, and 30). As there was not any solved 3D structure of fucosyltransferase belonging to the GT10 family at the time we found this motif, we decided to investigate it using the tools of molecular biology.

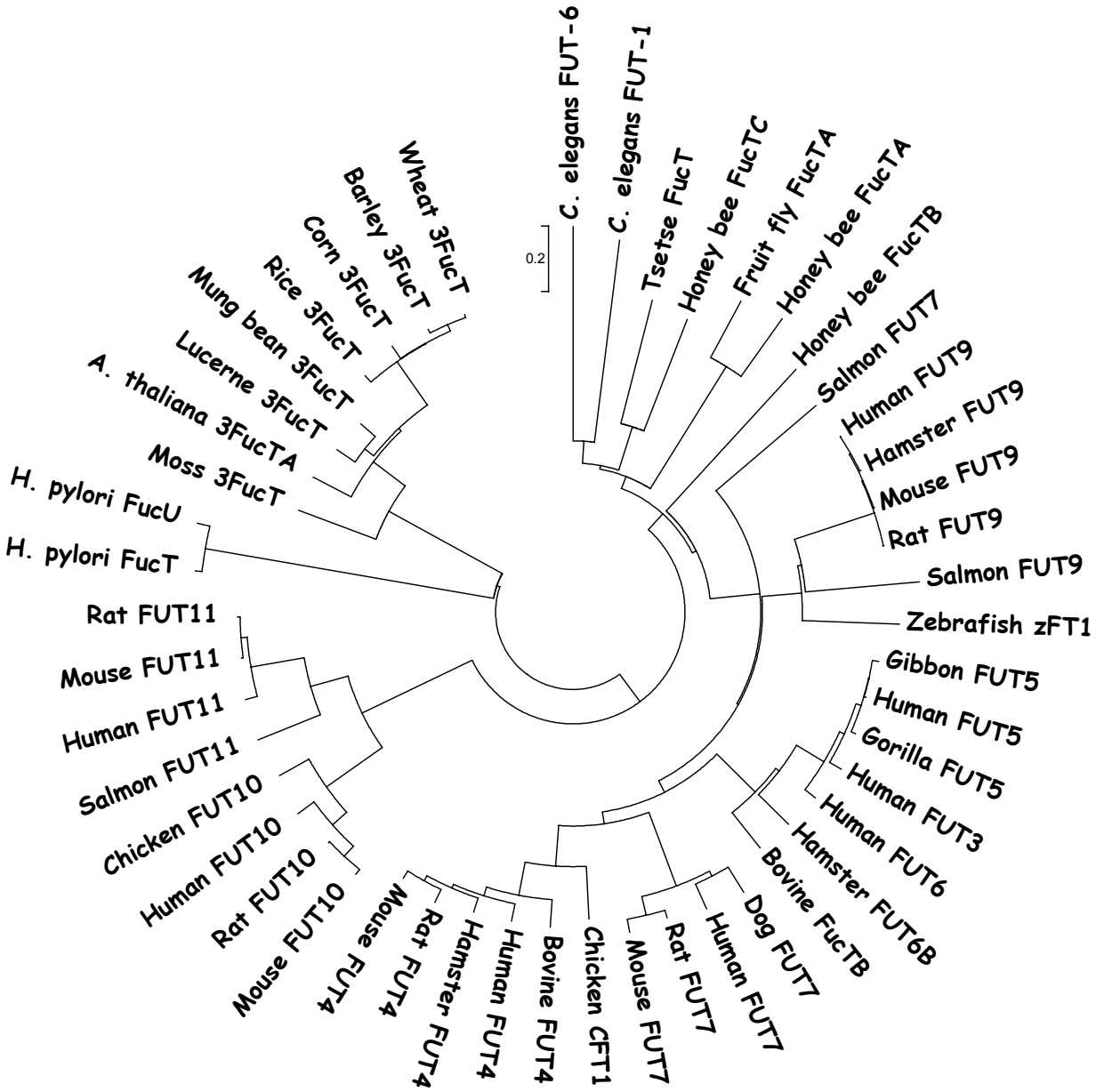


Fig. 19 **Phylogeny. Dendrogram comprising forty nine fucosyltransferases of the GT10 family.** As all investigated α1,3-fucosyltransferases of plant origin belong to Core type their names were changed to 3FucT in this picture (for UniProt codes and original names see chapter 2.1.)

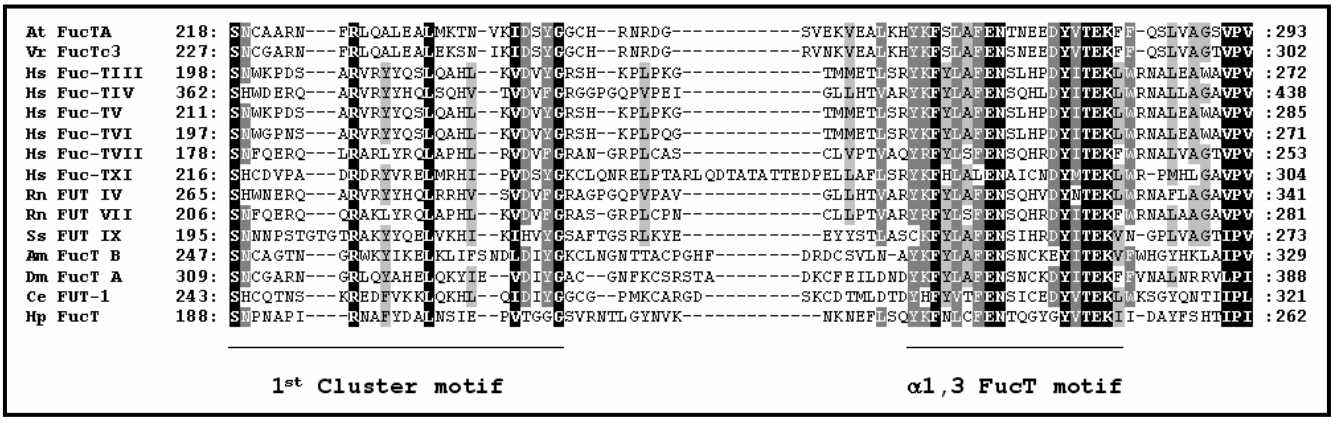


Fig. 20 **Alignment of peptide sequences of GT10 members of different origin depicting the 1st Cluster and α1,3 FucT motifs within the donor binding pocket.** Aligned sequences are as follows: *At FucTA* - Arabidopsis thaliana FucTA; *Vr FucTc3* - Vigna radiata FucTc3; *Hs Fuc-TIII* - Homo sapiens Fuc-TIII; *Hs Fuc-TIV* - Homo sapiens Fuc-TIV; *Hs Fuc-TV* - Homo sapiens Fuc-TV; *Hs Fuc-TVI* - Homo sapiens Fuc-TVI; *Hs Fuc-TVII* - Homo sapiens Fuc-TVII; *Hs Fuc-TXI* - Homo sapiens Fuc-TXI; *Rn FUT IV* - Rattus norvegicus FUT IV; *Rn FUT VII* - Rattus norvegicus FUT VII; *Ss FUT IX* - Salmo salar FUT IX; *Am FucT B* - Apis mellifera carnica FucT B; *Dm FucT A* - Drosophila melanogaster FucT A; *Ce FUT-1* - Caenorhabditis elegans FUT-1; *Hp FucT* - Helicobacter pylori FucT. Conserved residues of the 1st Cluster motif correspond to S218, R226, and G244 of *A.thaliana* FucTA.

Further, a CXXC motif, which was previously described in Lewis α3/4-FucTs, is conserved among all GT10 members of higher eukaryotes. This motif is localised relatively close to the C-terminus of all eukaryotic GT10 members except in plant core α1,3-fucosyltransferases where this motif is localised prior to the C-terminal region, which is longer (approximately 110 residues) than those of other GT10 eukaryotic members and is highly conserved among plant core α1,3-fucosyltransferases. Except for these Cys residues there are also some other conserved Cys residues in the sequences of human, plant, and nematode α1,3-fucosyltransferases. Beside the two conserved Cys residues of the CxxC motif, Holmes and co-workers found two other Cys residues, which are conserved in the sequences of human FUT III, V, and VI (these 4 Cys are also present in FUT IV, VII, and IX; all are members of the GT10 family) (Holmes EH *et al.*, 2000). They investigated possible disulphide bond formation between the conserved Cys residues in human FUT III, using a combination of proteolytic digestion and MS/MS analyses, which showed disulfide bridges between fist conserved Cys residue and the first Cys residue of the CXXC motif as well as between the second conserved Cys residue and the second Cys residue of the CXXC motif. Mutation of any of these conserved Cys residues in human FUT V led to a complete loss of enzyme activity. In the case of human FUT VII also another two Cys residues were found to be localised close to each other. In

the case of this enzyme all six Cys residues are involved in disulphide bridges. However, the bounding pattern is different from that of the human FUT III. In this case the disulphide bonds are formed in order first with second, third with fourth, and fifth with sixth (fifth and sixth are parts of the CXXC motif) (Holmes EH *et al.*, 2000). We also identified some other conserved Cys residues in the sequences of invertebrate α1,3-fucosyltransferases as well as among the sequences of plant core α1,3-fucosyltransferases, which might be involved in disulfide bond formation (Figure 22). The CXXC motif of *A. thaliana* FucTA comprises C364 and C367. The other Cys residues conserved among plant core α1,3-fucosyltransferases correspond to C91, C124, C128, C220, C246, C389, and C391 of *A. thaliana* FucTA.

As mentioned above, all plant core α1,3-fucosyltransferases possess a unique C-terminal region of approximately 110 residues, which is highly conserved and shows no sequence similarity to any known protein (Figures 21 and 24). Comparison with other GT10 protein sequences led us to propose the beginning of the C-terminal subdomain in the region 380-390, and particularly C389 was considered for experimental studies as being the first aa of this additional region in *A. thaliana* FucTA.

We did not find any conserved potential N-glycosylation site throughout of the protein sequences of GT10 family. However, there is one conserved putative N-glycosylation site common for all investigated plant core α1,3-fucosyltransferases, which corresponds to the Asn337 of *A. thaliana* FucTA. Furthermore, there is a conserved putative N-glycosylation site present in all *C. elegans* putative α1,3-fucosyltransferases (corresponding to Asn194 of *C. elegans* FUT-1) (Paschinger *et al.*, 2004).

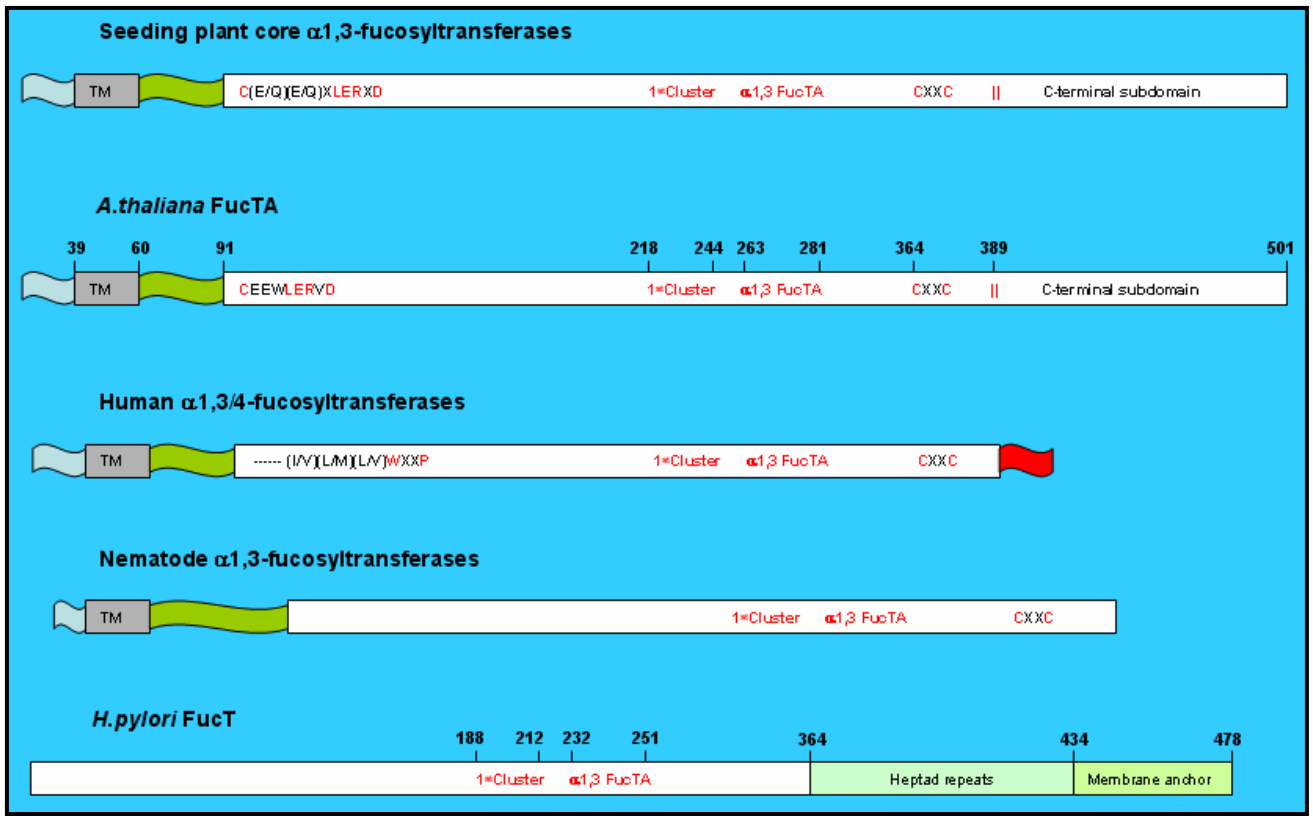


Fig. 21 **Schematic representations of some GT10 members of higher eukaryotes and *Helicobacter pylori* FucT.** Conserved motifs are represented by red letters. The first conserved motif among a particular group is listed at the beginning of the white bars (in the case there is such a sequence). Waves represent regions of variable lengths among a certain group; such as the stem region. Plant core α1,3-fucosyltransferases exhibit an extra C-terminal region, which does not show sequence similarity and functionality to any other known sequence. Heptad repeats of the *H.pylori* FucT are responsible for enzyme dimerisation.

Multiple sequence alignment of some plant core α1,3-fucosyltransferases, which prefer *GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc* as acceptor substrate, with the sequence of *C.elegans* FUT-1, which recognises *Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc* as acceptor substrate, revealed a region of approximately 20 residues prior to the conserved motifs towards the N-terminus of the investigated plant core α1,3-fucosyltransferases, which is not present in the FUT-1 sequence (Figures 22 and 23). We propose that this region could also account for donor substrate specificity.

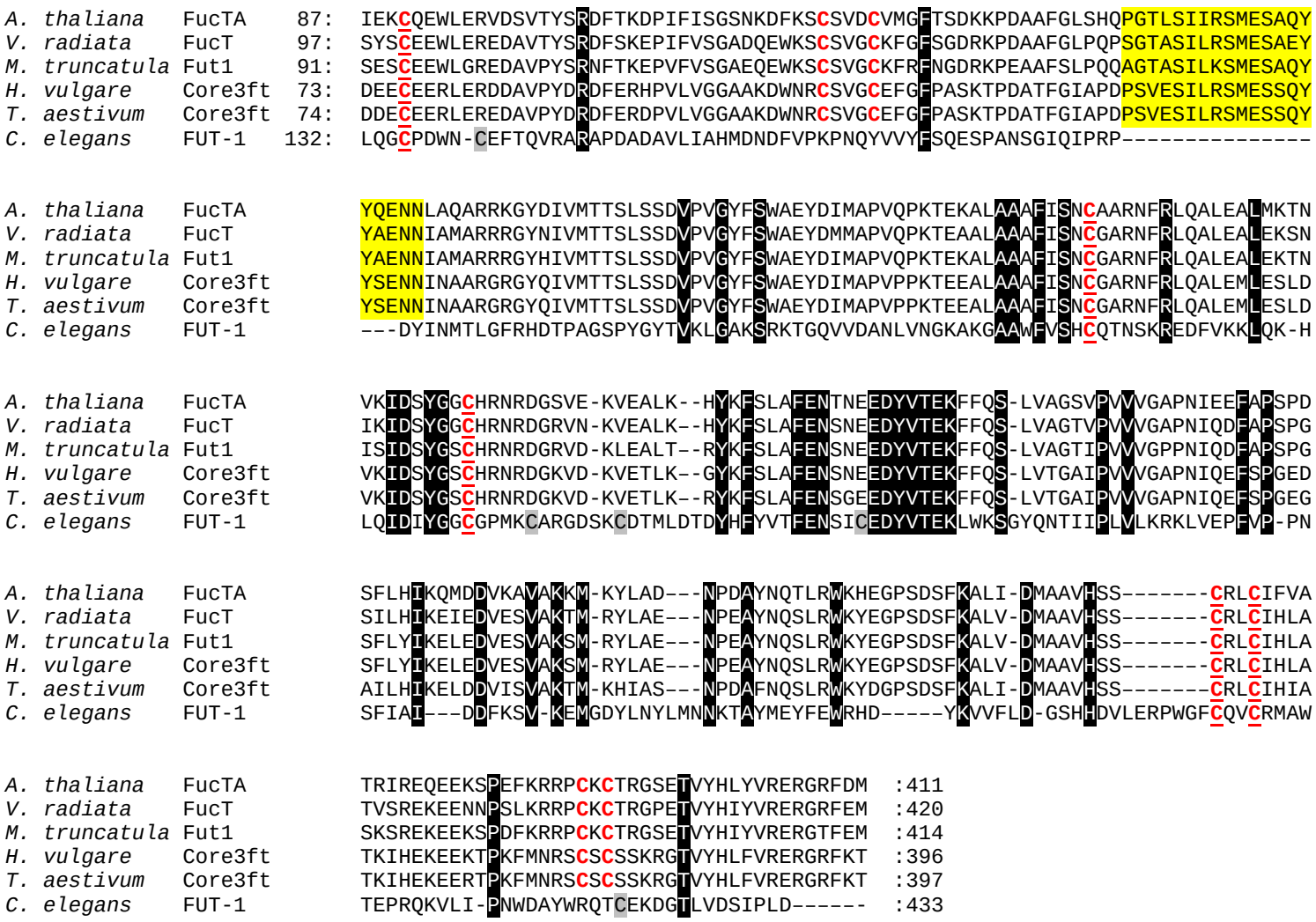


Fig. 22 **Section of sequence alignment of some plant core α1,3- fucosyltransferases and *C.elegans* FUT-1.** Cysteine residues, which are conserved only among plant core α1,3- fucosyltransferases are represented by red letters, while those conserved in both plant and nematode core α1,3-fucosyltransferases are represented by underlined red letters. The yellow frame represents a specific region of plant core α1,3- fucosyltransferases, which is missing in the sequence of the nematode core α1,3- fucosyltransferase. (*Medicago truncatula* Fut1, UniProt code: Q6Q8A9)

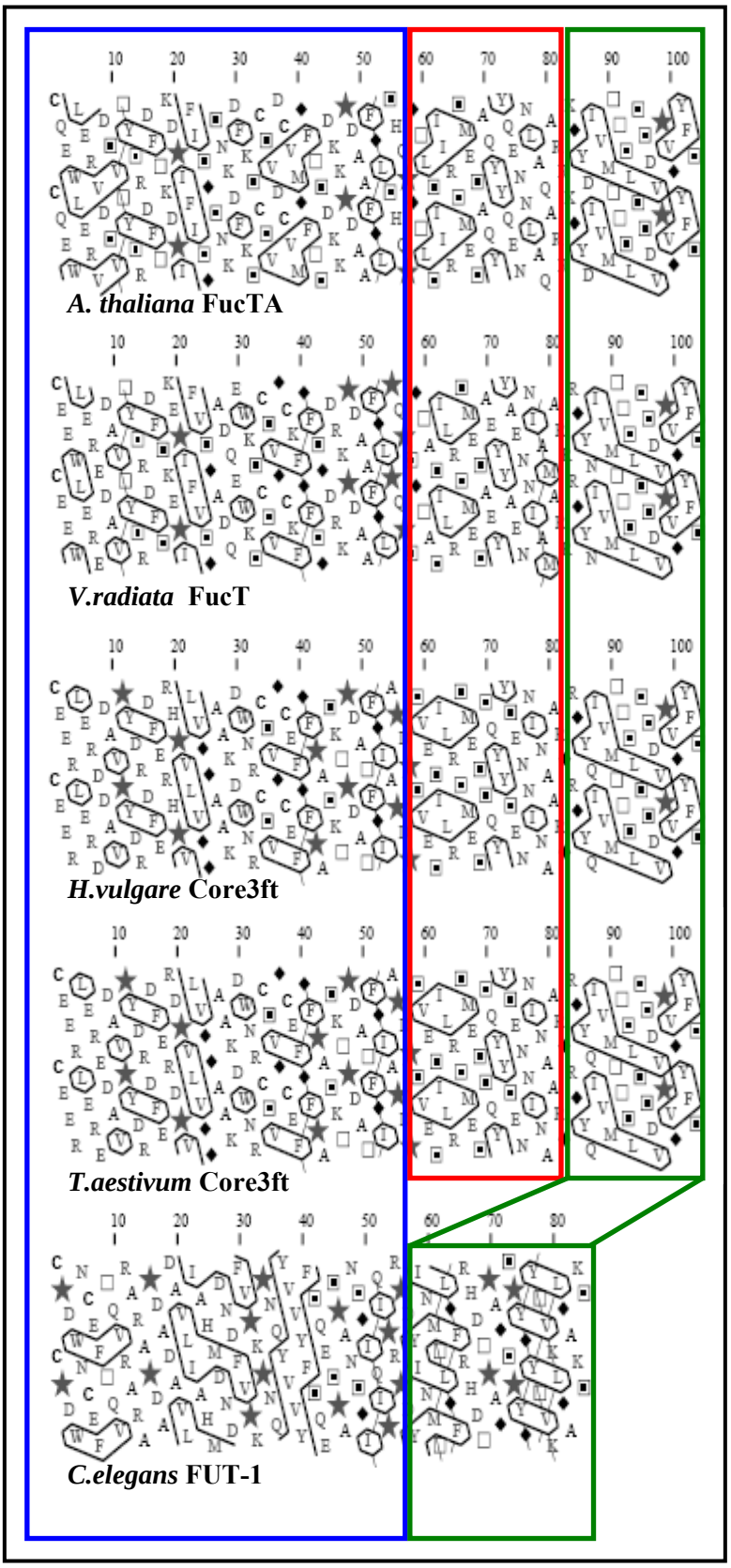


Fig. 23 **Alignment of HCA plots of corresponding regions of some plant core α1,3- fucosyltransferases and *C.elegans* core α1,3-fucosyltransferase sequences.** Sequences in the red frame represent the region missing in the nematode core α1,3-fucosyltransferase. Blue and green frames represent sequences surrounding the plant specific region.

As several divalent metal ions play an activator role in the activity of some α1,3-fucosyltransferases (Palma AS *et al.*, 2004; Murray BW *et al.*, 1997) we were looking for the canonical DXD motif in the protein sequences of GT10 members. This motif (or its variants as EXE, DXE, EXD) is known to be responsible for the donor substrate (nucleotide-sugar) binding in glycosyltransferases adopting GT-A fold (Breton C *et al.*, 1998; Wiggins CA *and* Munro S 1998). The two residues of this motif bind the donor substrate via coordination of a divalent metal ion, which interacts with the phosphate groups of the donor. This motif occurs in a short loop connecting two β-strands and is therefore easily identifiable using HCA. However, this inspection did not show such a conserved motif among members of the GT10 family.

```
A. thaliana      FucTA     375:  REQEEKSPEFKRRPCKCTRGSE-TVYHLYVRERGRFDMESIFLKDGNLTLEALESAVLAKFMSLRYEPIWKKERPASLRGDGK
V. radiata       FucT      384:  REKEENNPSLKRRPCKCTRGPE-TVYHIYVRERGRFEMESIYLRSSNLTLNAVKAAVVLKFTSLNLVPVWKTERPEVIRGGSA
M. sativa        FUT11     379:  REKEEKSPDFKKRPCKCTRGSE-TVYHIYVRERGTFEMESIYLRSSNLTLESFKTAVLTKFTSLNHVPVWKPERPQILKGGDK
H. vulgare       Core3ft   360:  HEKEEKTPKFMNRSCSCSSKRG-TVYHLFVRERGRFKTENIYLRSDQLTLGALKSAVHDKFSSLKHVPIWKDERPSSIRGGDE
T. aestivum      Core3ft   361:  HEKEERTPKFMNRSCSCSSKRG-TVYHLFVRERGRFKTESIYLRSDQLTLGALESAVHGKFRSLKHVPVWKDERPSSIRGGDE
O. sativa        FUT11     387:  HEKEERTPKFMNRPCSCSSKRG-KVYHLFVRERGRFKTESIFLRSDQLTMGALESAVLAKFRSLNHVPVWKDERPPSIRGGDE
Z. mays          FucT      372:  HLKEERTPKFTNRPCSCSTKKG-TIYHLFIRERGRFKSESIYMRSGQLTLGALESAVLGKFRSLNHVPVWKDERPPSIRGGDD
Ph. Patens       Fuct      408:  RLKEEAAAP--KRPCKCTSKSGSTLYHLYVRERGRFEMESVFIEGSKLSLAHLKQVVVDKFTALKHGPIWKTERPGVIRGNSD
H. sapiens       FUT3      349:  RYQTVRSIAAWFT-------------------------------------------------------------------
H. sapiens       FUT5      362:  RYQTVRSIAAWFT-------------------------------------------------------------------
H. sapiens       FUT6      348:  RYQT-RGIAAWFT-------------------------------------------------------------------
C. elegans       FUT-1     404:  RQKVLIPNWDAYWRQTCEKDGTLVDSIPLD-------------------------------------------------
D. melanogaster  FucTA     462:  QLRKPRWYTDLNDWWRGPGVCTTRSWRNFKARKDVISDSSDD----------------------------------------


A. thaliana      FucTA           LRVHGIYPIGLTQRQALYNFKFEGNSSLSTHIQRNPCPKFEVVFV  :501
V. radiata       FucT            LKLYKIYPIGLTQRQALYTFSFKGDADFRSHLENNPYAKFEVIFV  :510
M. sativa        FUT11           LKVYKIIPAGLTQRQALYTFQFNGDVDFRSHLESNPCAKFEVIFV  :505
H. vulgare       Core3ft         LKVYKIYPIGLTERQALYKFQFSDDAEVARYIKGHPCAKLEVIFV  :486
T. aestivum      Core3ft         LKVYKIYPIGLTERQALYKFQFSDDAEVARYIKGHPCAKLEVIFV  :487
O. sativa        FUT11           LKVYKIYPIGLTQRQALYQFRFRDDADLDKYIKDHPCAKLEVIFV  :513
Z. mays          FucT            LKLYRIYPVGLTQRQALYGFRFRDDSELEQYIKDHPCAKLEVIFV  :498
Ph. Patens       Fuct            LRIYKIYPVGLTQREALYTWDFGGDKGLKAMVQKQPCLQLEVVFV  :533
H. sapiens       FUT3            ----------------------------------------------  :361
H. sapiens       FUT5            ----------------------------------------------  :374
H. sapiens       FUT6            ----------------------------------------------  :359
C. elegans       FUT-1           ----------------------------------------------  :433
D. melanogaster  FucTA           ----------------------------------------------  :503
```

Fig. 24 **Sequence alignment of the C-terminal regions of plant core α1,3- fucosyltransferases, selected human α1,3- fucosyltransferases, and two invertebrate core α1,3- fucosyltransferases.**

3.1.2. Homology modelling

The recently solved crystallographic structure of FucT from *Helicobacter pylori* (Sun HY *et al*., 2007) obtained from protein crystals soaked with the donor substrate (GDP-Fuc) is consistent with our prediction that the 1st Cluster motif, the α1,3 FucT motif, and the sequence lying between them, all together form the donor substrate binding pocket (Figures 20, 27, and 28). Moreover, it also proved our prediction that the conserved serine and arginine of the 1st Cluster motif form hydrogen bonds with the donor substrate (Figure 25).

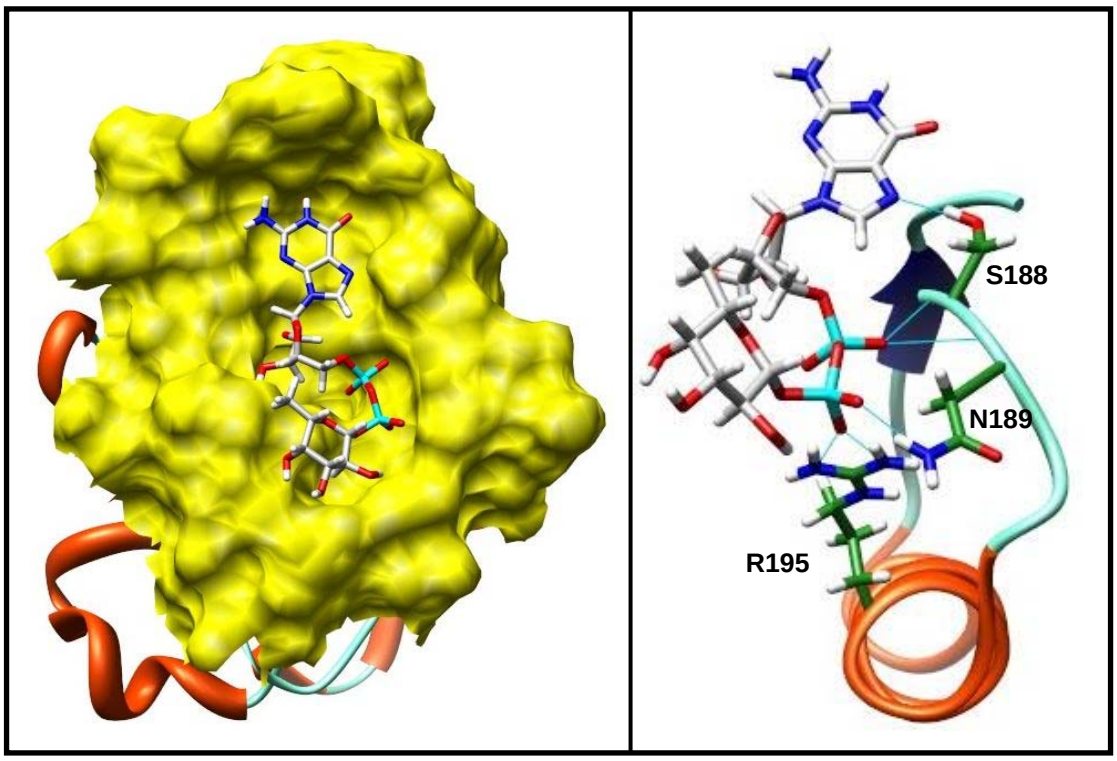


Fig. 25 **Structure of the donor binding pocket of FucT from *Helicobacter pylori* and detail of the 1st Cluster motif** . PDB data of FucT soaked with GDP-Fucose (PDB code: 2NZY) - were used as source and visualised by UCSF Chimera.

As a template for homology modelling we used the structure of FucT from *Helicobacter pylori* (Sun HY *et al*., 2007). First of all a good protein alignment was needed to model the donor binding pocket of the *Arabidopsis thaliana* FucTA. The most conserved regions of *H.pylori* FucT and *A. thaliana* FucTA were therefore aligned. Several alignments of these sequences were used to build models by Modeller9v1 (Eswar N *et al.*, 2007). The best backbone overlapping (Figure 26) was obtained using alignment shown in figure 27C. The obtained model was then visualised and protonated using UCSF Chimera program (Sanner MF *et al.*, 1996; Pettersen EF *et al.*, 2004; Goddard TD *et al.*, 2005) in the presence of the donor substrate. Data for GDP-Fucose were retrieved from the PDB: 2NZY structure. Charge -2 was assigned to the donor substrate and a subsequent energy minimisation was performed using the tool ANTECHAMBER of UCSF Chimera (Wang J *et al.*, 2006).

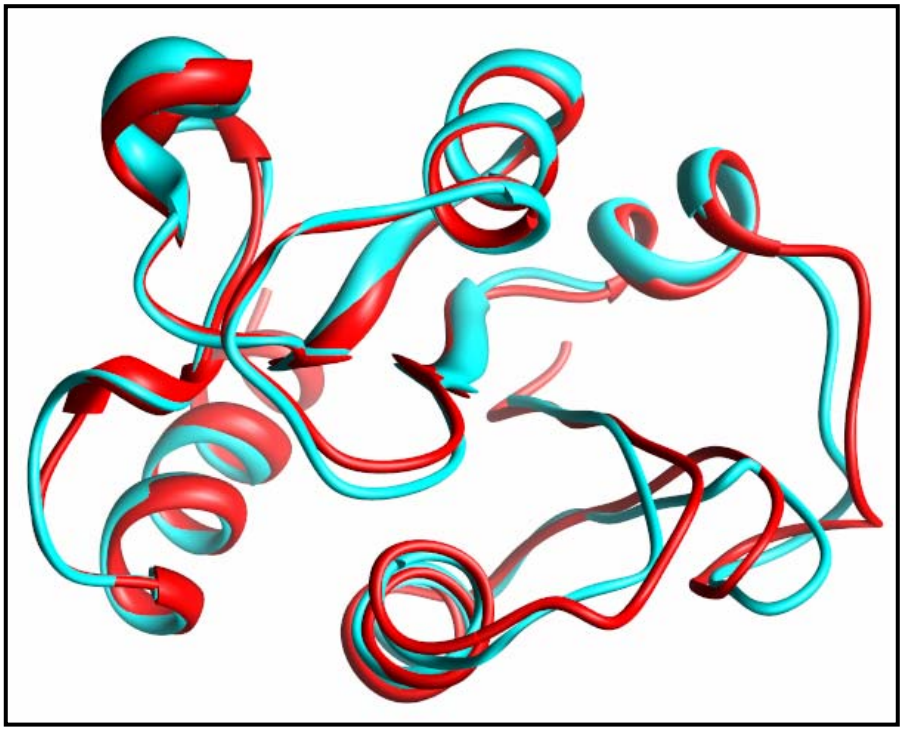


Fig. 26 **Backbone overlapping of the donor binding pocket of *H. pylori* FucT (in red) and that of the *A. thaliana* FucTA model (in cyan).** The *A. thaliana* FucTA model presented here is before energy minimisation.

All of the important amino acid residues, which form hydrogen bonds with the donor substrate in the template are present in the model too, and are almost in the same positions. Moreover, in our model one can see how an additional hydrogen bond is formed between the hydroxyl-group of Ser253 and the guanine of the donor substrate as well as a stacking effect between a Tyr243 and the nitrogen 7 of the guanine (Figures 27 and 28).

We also tried to find structural templates to either model the whole enzyme or at least its N-terminal part which is supposed to be responsible for the acceptor substrate binding. A BLAST search performed on PDB using the N-terminal region (C91 – L212) of FucTA did not give any usable match for homology modelling. We also tried to align this region by Clustal W (using GONNET weight matrix) with the N-terminal domain of the *H. pylori* FucT as well as with the N- and C-terminal domains of human FUT8 (the latter enzyme recognises the same acceptor substrate as FucTA). Unfortunately, none of the models built showed any structural match.

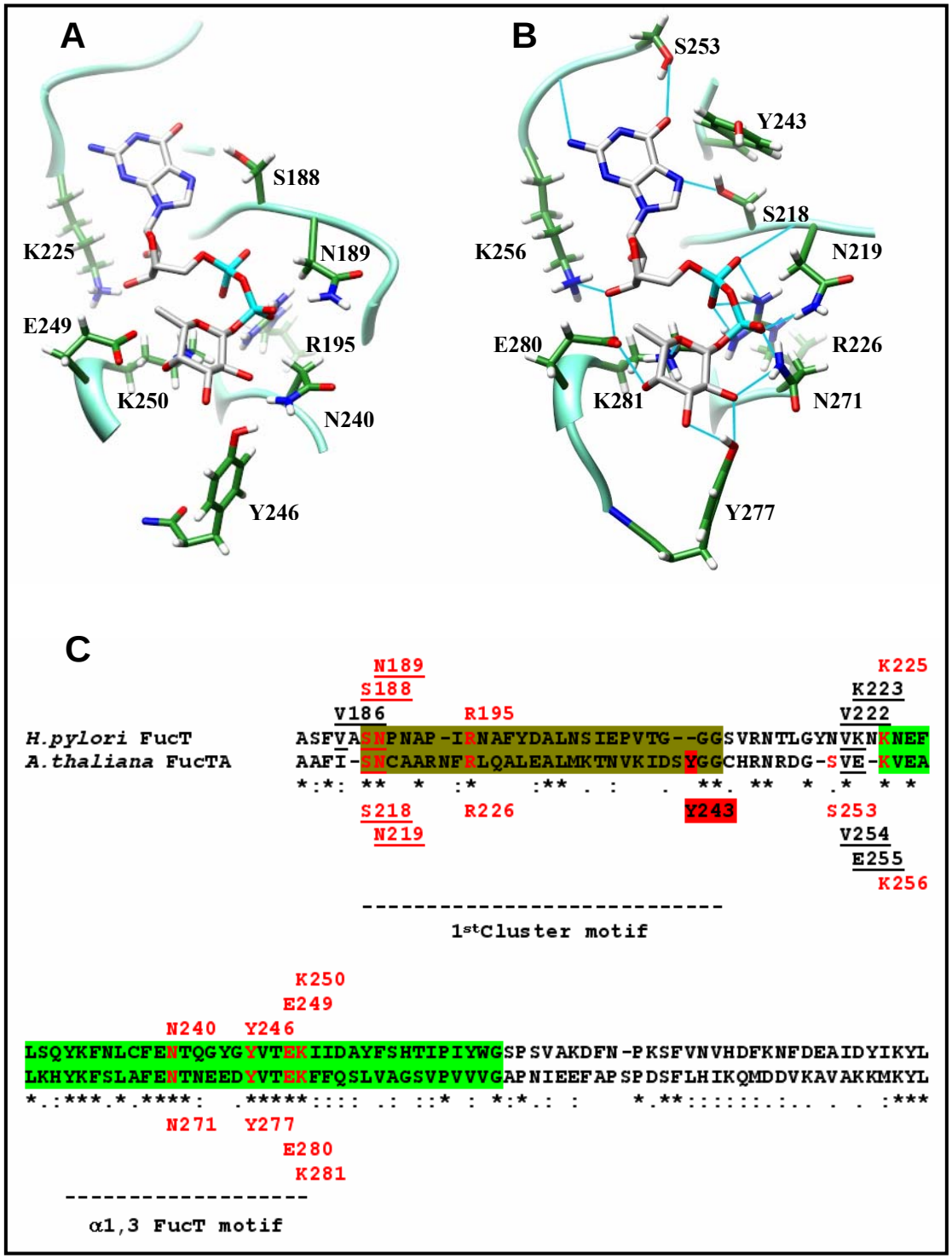


Fig. 27 **Environment of the donor substrate GDP-Fuc in (A) the structure of *H.pylori* FucT and (B) in our proposed model of the donor binding pocket of *A.thaliana* FucTA. (C) Alignment of *H.pylori* FucT and *A.thaliana* FucTA protein sequences used to build the 3D model of the donor binding pocket for FucTA.** Red letters represent residues forming hydrogen bonds with the donor substrate. Y243 of the 1st Cluster motif of FucTA is proposed to be involved in a stacking effect between its aromatic ring and nitrogen 7 of the guanine. Underlined letters represent possible positions for hydrogen bond forming between the backbone (peptide chain) and the donor substrate.

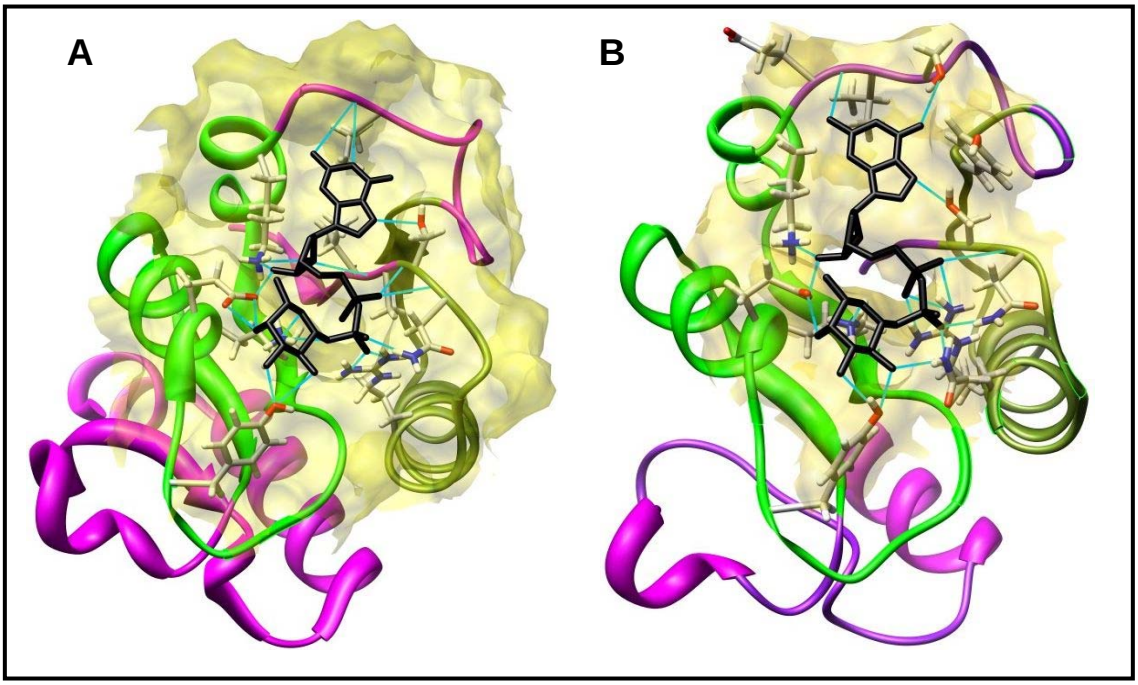


Fig. 28 **Ribbon structures and surfaces of donor substrate binding pockets in the proximity of the donor substrate**. **A: Donor binding pocket of *H.pylori* FucT.** Visualised upon PDB data (PDB code: 2NZY). **B: 3D model of *A.thaliana* FucTA donor binding pocket.** (Ribbon colours correspond to regions marked with same colors in Figure 15 C, all other regions are in magneta)

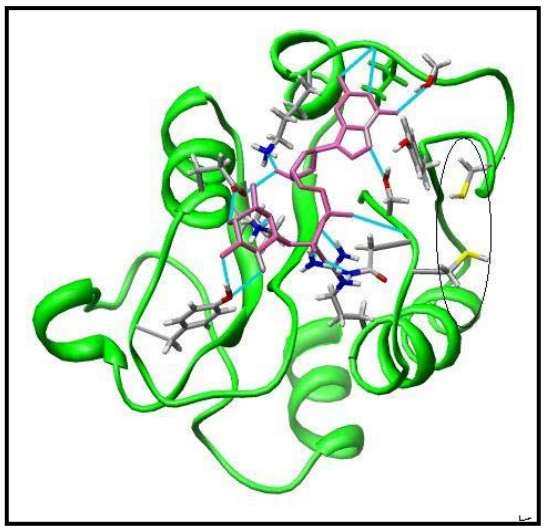


Fig. 29 **Two proximal Cys residues of the 1st Cluster motif and the loop of the donor substrate binding pocket viewed in the model of *A.thaliana* FucTA.** The distance between the sulphur atoms of C220 and C246 is 3.98 Å.

Examination of the 3D model revealed that the two cysteine residues (C220 and C246) of the 1st Cluster motif in AtFucTA (absent in HpFucT), are close enough to form a disulfide bridge. Such a bridge is expected to modify the shape of the loop lying between the 1st Cluster and α1,3 FucT motifs (Figure 29).

3.2. Investigation and engineering of *Arabidopsis thaliana* FucTA

3.2.1. Truncated and immuno- and/or affinity-tagged fusion mutants of *A. thaliana* FucTA

In order to determine the minimal catalytic domain of *A. thaliana* core α1,3-fucosyltransferase FucTA, some N-terminal truncation mutants had been previously prepared. The first catalytically active construct of FucTA, which was successfully expressed in *P. pastoris* is lacking 66 residues (Δ1-66). Another mutant lacking 88 residues (Δ1-88) was still active, while a mutant lacking 95 residues (Δ1-95) lost its activity (Bencúrová M *et al.*, 2003).

Since one of our goals was to study the effect of N-glycosylation on the FucTA activity we were looking for an expression system which provides protein N-glycosylation. For our previous study of the *Vigna radiata* FucTc3 (which is not part of the work presented here), we chose an insect cell (Sf9/baculovirus) expression system. Though this expression system gave an even and well defined glycosylation of the recombinant FucTc3, the level of protein expression was very unsatisfactory. Moreover, the heterologous expression of FucTA was successful in *P. pastoris*, which made this expression system our preferential choice.

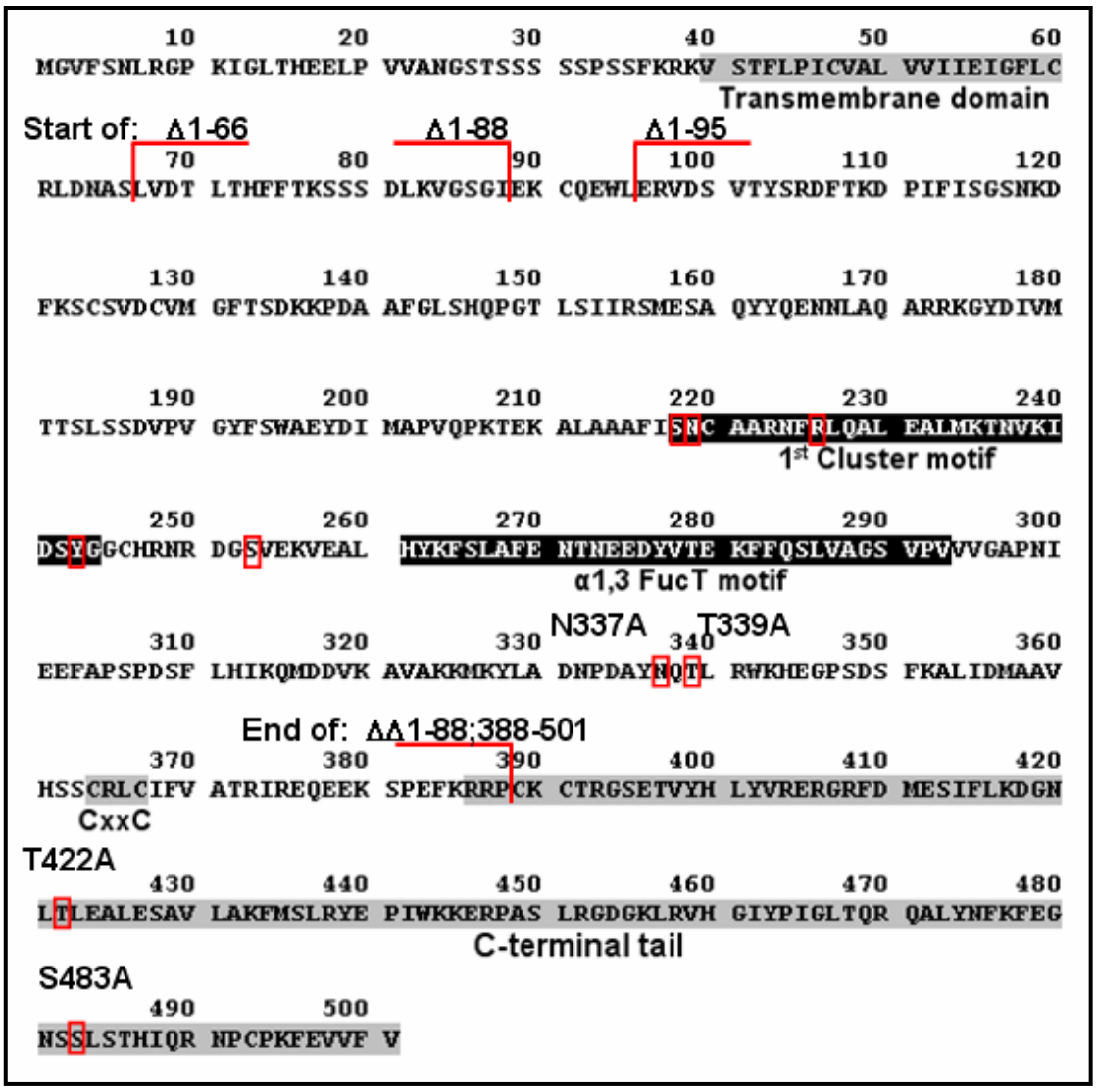


Fig. 30 **Protein sequence of *Arabidopsis thaliana* FucTA,** showing positions of the beginning and ends of the truncated constructs (Δ1-66, HFΔ1-88, Δ1-95, HFΔΔ1-88;388-501) as well as positions of single point mutations (S218A, N219A, R226A, Y243A, S253A, N337A, T339A, T422A, S483A) applied in this work. Shaded sequences represent conserved parts among GT10 members. The predicted transmembrane domain, the CxxC motif, and the proposed C-terminal tail unique to plant core α1,3-fucosyltransferases, the 1st Cluster motif and α1,3 FucT motif (an extended version based on our outputs from bioinformatics analysis), which are present in protein sequences of all known GT10 members (marked by white letters in black) are presented. The donor binding pocket of our model comprises the sequence between A214 and L329.

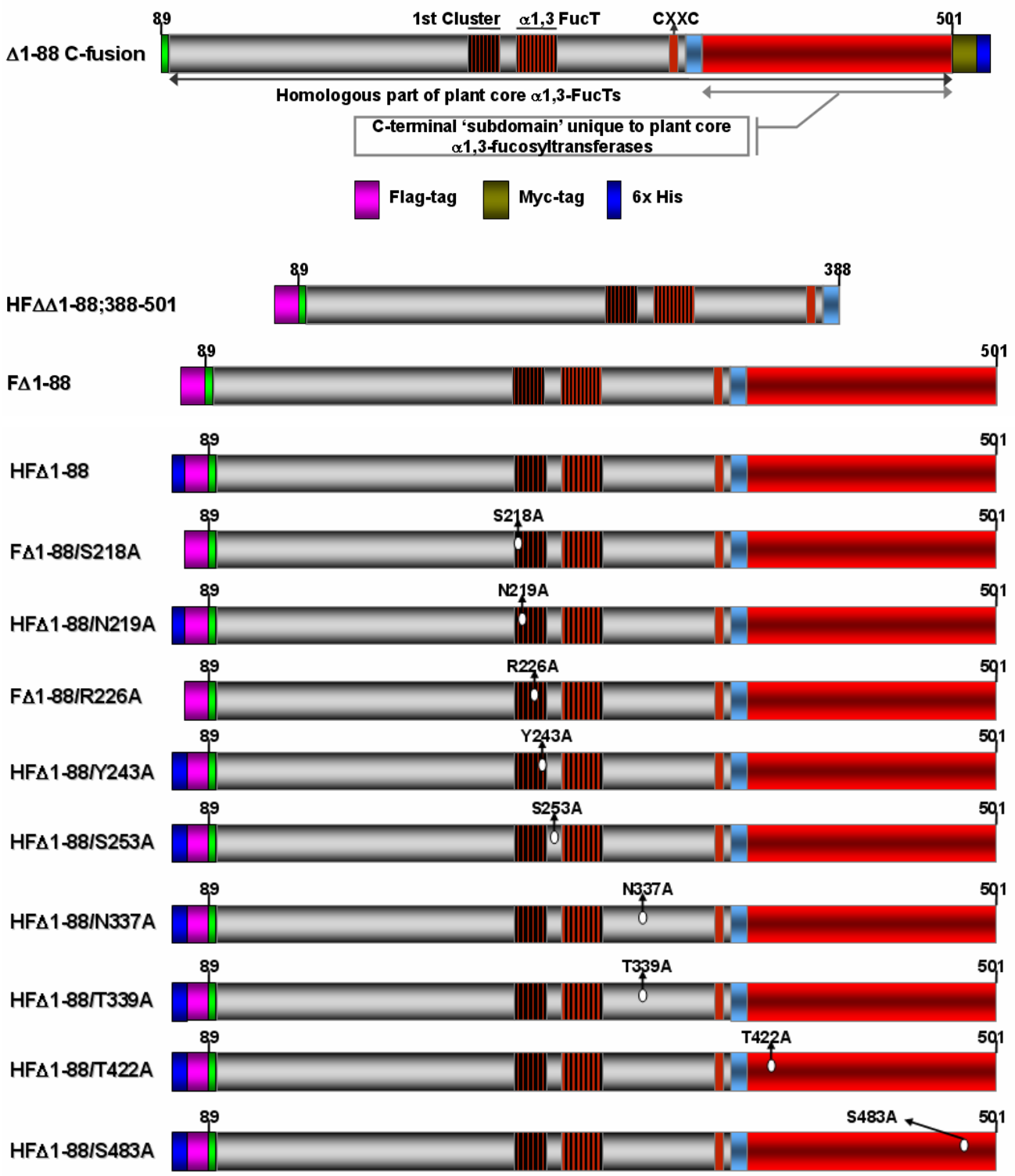


Fig. 31 **Schematic representations of all recombinant *A. thaliana* FucTA forms prepared for the presented work.**

Each protein construct, which is subject of this work, is represented in figure 31 for better understanding.

We initially considered of use a Myc-tag followed by 6xHis C-terminal fusion in the case of the Δ1-88 mutant to facilitate its purification. In fact, this C-terminal peptide fusion resulted in lack of protein expression (in *Pichia pastoris*). Thus, in further experiments we engineered N-terminal fusion of Δ1-88 mutant with immunogenic Flag-tag (home made vector pPICZαFlagC3 see appendix; result of engineering FΔ1-88). This mutant was fully active and specifically visualized by immunobloting (Figure 32 C). In a second step, in order to enable purification of recombinant FucTA mutants from the supernatant, N-terminal fusions of these mutants with tandem of 6 His residues followed by Flag-tag were prepared (home made vector pPICZαHisFlag, see appendix). This modification to the N-terminus did not affect the enzyme activity (Figure 32 B). Relative concentrations of recombinant enzymes were estimated using a calibration curve of ELISA by using of Flag-BaP (see Material and methods; anti-Flag ELISA).

Furthermore, in an initial experiment aimed at defining whether the last 113 residues delineating the C-terminal subdomain unique to plant core α1,3-fucosyltransferases are necessary for the enzyme activity. We prepared a truncation of the C-terminal subdomain derived from the mutant HFΔ1-88 (HFΔΔ1-88; 388-501 mutant; Figure 32 C). This double truncation did not affect recombinant protein expression, however led to dramatic decrease of activity, less than <1% (Figure 32 B). On the other hand, the K_m value of this mutant for GDP-Fuc was determined to be 0.1 mM (this value is the same as for the HFΔ1-88; see next section), which suggests that this part of the plant core α1,3-fucosyltransferases will rather be involved in stabilisation of the enzyme's structure and maybe in some other, as yet unknown, functions rather than to be involved in donor substrate binding.

In order to understand structure-function relationships of α1,3-fucosyltransferases as well as to prepare the basis for protein crystallisation trials, many studies have described deletion mutants of enzymes of different origin. From the previous results mentioned above (Bencúrová M *et al.*, 2003), it is evident that the minimal catalytic domain of the *A. thaliana* FucTA starts between E89 and E96. The first conserved motif among seed plant core α1,3-fucosyltransferases **C**(E/Q)(E/Q)(W/R)**LE** is located between these residues and starts with C91. We propose this motif to be the determinant of the beginning of minimal catalytic domains of seed plant core α1,3-fucosyltransferases.

Our results are consistent with GT10 members of higher eukaryotes being sensitive to protein manipulation on their C-terminus, while the CTS domain is not essential for the enzyme activity and can be replaced by other peptide sequences (like immuno- and/or affinity-tags).

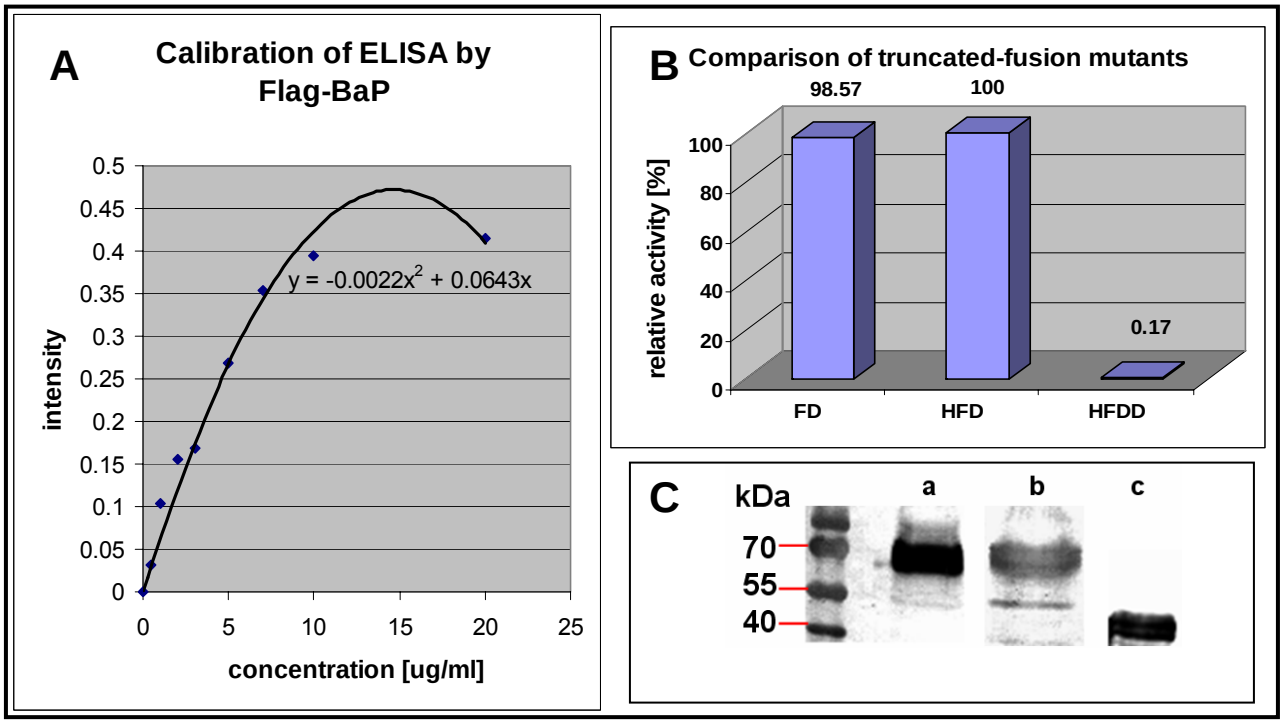
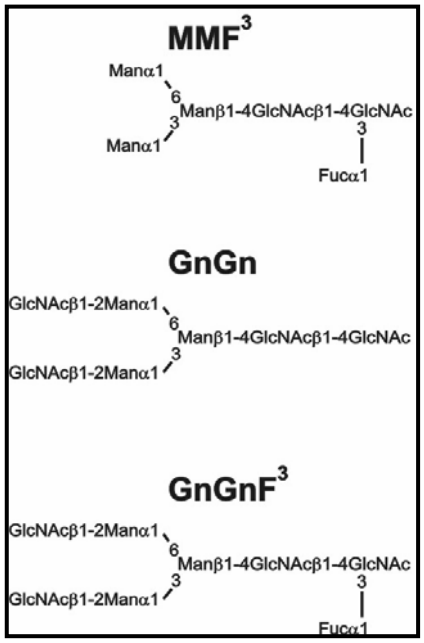


Fig. 32 **A: Calibration curve of ELISA by using of Flag-BaP.** Performed with anti-Flag monoclonal antibodies from mouse (M2) from Sigma, diluted 1:5000 (see Material and methods; anti-Flag ELISA). To calculate relative concentration of proteins the first "half" of the quadratic equation was taken (Samples were analyzed with a SLT Spectra plate reader, using the machine's PNP program to obtain absorbance values). **B: Comparison of relative specific activities of hybrid and double truncated FucTA constructs**. Activity of the HFΔ1-88 was used as reference. Enzyme activities were evaluated by reverse phase HPLC-assisted measurement of the conversion of $GnGnF^6$-NST-dansyl (see Scheme 1). Protein concentrations (as all constructs bear Flag-tag) were estimated by anti-Flag ELISA. (*FD*: FΔ1-88; *HFD*: HFΔ1-88; *HFDD*: HFΔΔ1-88;388-501).**C: Western blotting** with mouse anti-Flag antibodies (1:5000). FucTA proteins were expressed and purified with the same conditions (see Material and methods; HisTrap purification) except for FΔ1-88, which was purified by Affigel Blue affinity chromatography. *Line a:* FΔ1-88; *line b:* HFΔ1-88; *line c:* HFΔΔ1-88;388-501.

3.2.2. Biochemical characterisation of the reference hybrid protein HisFlagΔ1-88 FucTA (HFΔ1-88)

Previous studies have shown that FucTA enzyme lacking 66 residues from N-terminus indeed catalyse synthesis of anti-HRP epitope *in vitro* (Wilson *et al.*, 2001; Bencúrová M *et al.*, 2003). To demonstrate that the enzyme is responsible for the formation of a core α1,3 linkage between the fucose and the acceptor substrate, a neoglycoconjugate of blood transferrin carrying a GnGn-glycopeptide (Scheme 1) was incubated with purified HFΔ1-88 in the absence and presence of GDP-Fuc. Dot-blot using BioTrace™ NT Nitrocellulose Transfer Membrane (Pall Corporation) was employed. Samples of reaction mixture were spoted on the membrane which was then treated by anti-HRP antibodies and processed as it is described in section 2.11. of Materials and methods. As predicted from the other substrate specificity experiments (Wilson *et al.*, 2001; Bencúrová M *et al.*, 2003), the thereby modified GnGn-transferrin conjugate was recognized by anti-HRP (Figure 33 D). Since there was not available a dansylated GnGn substrate, for quantitative enzyme assays dansyl-modified form of a $GnGnF^6$ glycopeptide (which is also recognised as acceptor substrate by FucTA), $GnGnF^6$-NST-dansyl, was used as acceptor substrate. Enzymatic assay mixtures analyzed by reversed-phase HPLC using MZ Analysentechnik Hypersil column (250 x 4.0 mm; ODS C18 5 μm) at isocratic 15 minutes run (flow 1.5 ml/min) with 8.5 % of solvent B and 91.5 % solvent A (solvent B: 95% acetonitrile in water; Solvent A: 0.05% TFA in water). Excitation λ was set on 315 nm and emission λ on 550 nm.



Scheme 1 **Scheme of oligosaccharide structures used in dot-blot test to prove formation of the anti-HRP epitope *in vitro.***

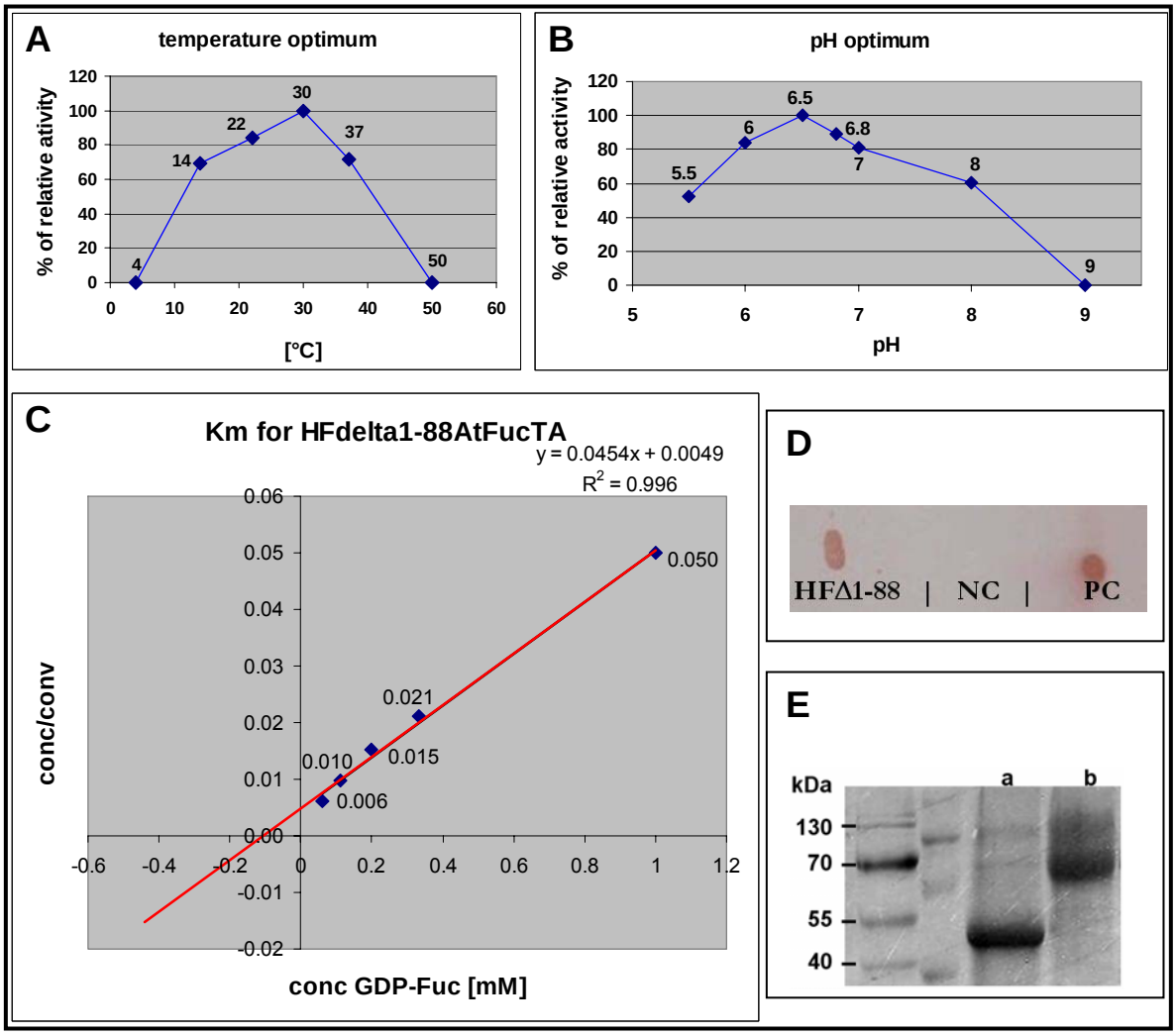


Fig. 33 **Characterisation of FucTA (HFΔ1-88). A: Temperature optimum of FucTA (HFΔ1-88)** at pH 6.5 (see Material and method; *A. thaliana* FucTA enzyme assays). **B: pH optimum of FucTA (HFΔ1-88)** at 30°C (see Material and method; *A. thaliana* FucTA enzyme assays). **C: Determination of FucTA (HFΔ1-88) K_m value for GDP-Fuc** by Hanes-Woolf plot (see Material and method; *A. thaliana* FucTA enzyme assays). **D: Recombinant FucTA catalyse formation of the anti-HRP epitope *in vitro*.** Purified HFΔ1-88 was incubated with GnGn-transferrin (see Scheme 1) in the presence and absence of GDP-Fuc (see Material and method; *A. thaliana* FucTA enzyme assays). The incubation mixture was then subjected to Dot blotting using rabbit anti-HRP as primary antibody (*spots HFΔ1-88*: partially purified HFΔ1-88; *spot NC*: partially purified HFΔ1-88 incubated without GDP-Fuc served as negative control; *spot PC*: MMF^3-BSA (see Scheme 1) as positive control). **E: PNGase F treatment of FucTA (HFΔ1-88).** HFΔ1-88 purified by HisTrap affinity chromatography was treated by PNGase F and analyzed by Western blotting (*line a*: PNGase F treated HFΔ1-88; *line b*: purified HFΔ1-88). For conditions see Material and methods (PNGase F treatment of *A.thaliana* FucTA; Western blotting)

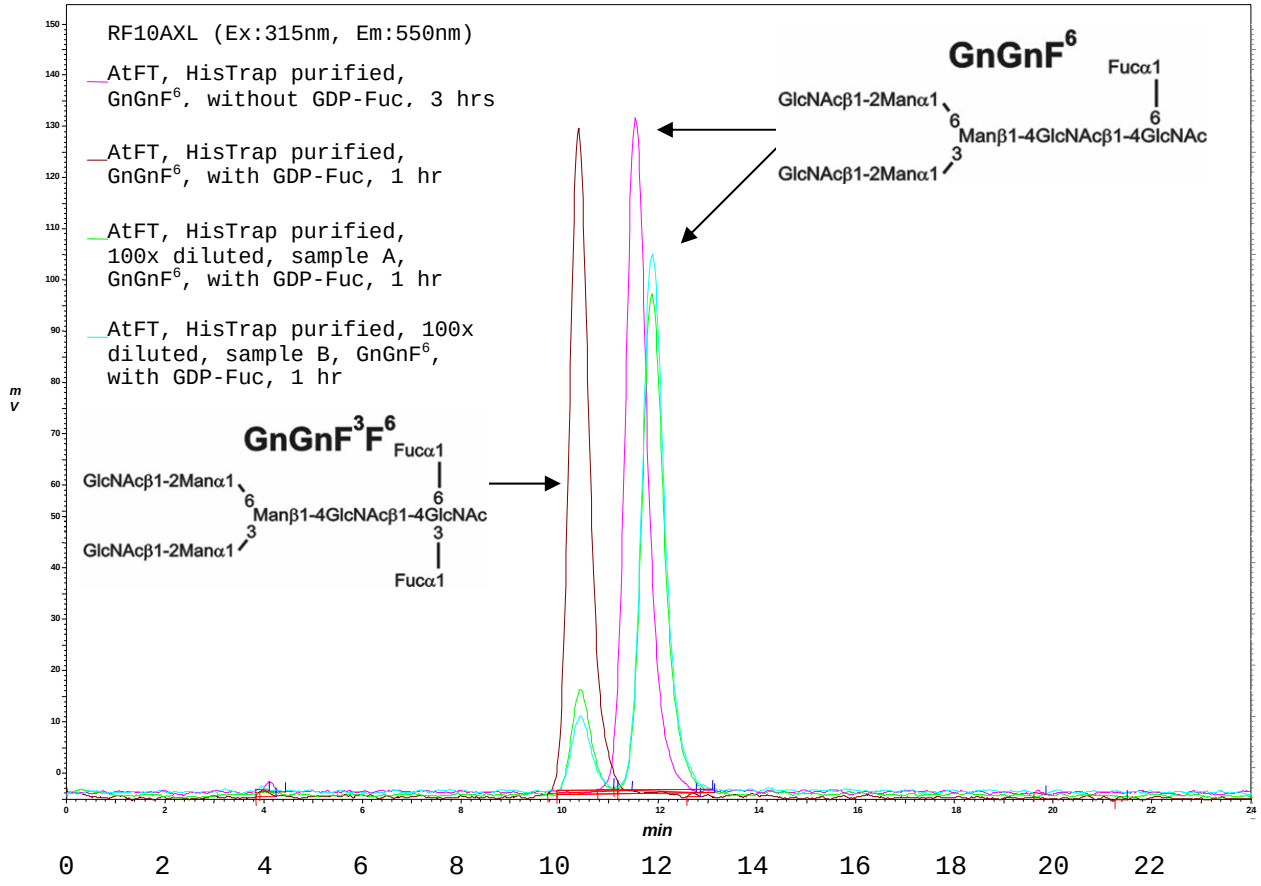


Fig. 34 **Reverse-phase HPLC analysis of reaction product of FucTA (HFΔ1-88) activity.** After incubation of 1 µl of either concentrated or diluted HisTrap purified HFΔ1-88 with GnGnF^{6}-NST-dansyl and with or without GDP-Fuc for 1 h at 30 °C, the assay mixtures were analysed by reversed-phase HPLC and fluorescence detection (*Materials and Methods*; Protein purification; *A. thaliana* FucTA enzyme assays).

MALDI-TOF MS analysis of another conjugate GnGnF^{6}-NST-dansyl (Figure 34), which we incubated in the presence of GDP-Fuc showed the presence of a peak with a composition consistent with a GnGnFF structure, whereas this peak was absent in the control (without addition of GDP-Fuc). MALDI-TOF MS/MS analysis of the reversed-phase HPLC peak, which was supposed to correspond to the GnGnF^{3}F^{6}-NST-dansyl product (Figure 34) of the enzyme activity of HFΔ1-88, also proved the presence of a GnGnFF structure.

Biochemical characteristics of the fusion enzyme HFΔ1-88 were determined: The optimal pH for the reaction (measured at 30°C) was found to range between pH 6 and 7 with a maximal value at pH 6.5.

The K_m value of HFΔ1-88 for the donor substrate was determined by a series of assays in the presence of decreasing concentration (1 mM – 0.06 mM) of GDP-Fuc, using $GnGnF^6$-NST-dansyl as acceptor substrate at a fixed concentration 2 μM (data was interpreted by Hanes-Woolf plot; Figure 33). The K_m value for GDP-Fuc was determined to be 0.1 mM.

Several eukaryotic members of the GT10 family have been biochemically characterised over the years. The pH optimum for Golgi-type members of GT10 family such as FucT (a Lewis-type fucosyltransferase) from *Vaccinium myrtillus* (bilberry) is pH 7.0 (Palma AS *et al.*, 2001), human Fuc-TVII encoded by the FUT7 gene is pH 7.5, and human Fuc-TV encoded by the FUT5 gene is between pH 5.0 and 7.0 (Shinoda K *et al*., 1997; Holmes EH *et al.*, 1995). The pH optimum of *C.elegans* FUT-1 is between pH 6.5 and 8 with maximal activity at pH 7.5 (Paschinger K *et al.*, 2004). Therefore the pH optimum of *Arabidopsis thaliana* FucTA is within the usual range of Golgi-type α1,3-/1,4-fucosyltransferases. Since the Golgi complex displays pH around 6.5 in different organisms (Seksek O *et al.*, 1995; Kim JH *et al.*, 1996; Llopis J *et al.*, 1998) such pH optima are consistent with the intracellular milieu of these enzymes.

In the case of the donor substrate the K_m value was determined to be 0.1 mM for *C.elegans* FUT-1 (Wilson IB, unpublished data), compared to K_m=0.07 mM for *Heliobacter pylori* FucT (Ma *et al.*, 2006), K_m=0.02 mM for human FUT VI (Jost F *et al.*, 2005), and K_m=16.4 μM for human FUT VII (Shinoda K *et al*., 1997). The K_m value of *A.thaliana* FucTA is the same as that of *C.elegans* FUT-1 and rather comparable to that of *Helicobacter pylori* FucT.

Divalent metal ion dependency of HFΔ1-88 was examined on enzyme preparation deprived of divalent metal ions by dialysis (against weak 25 mM Tris-HCl pH7.5 buffer) in the absence and presence of 10 mM divalent metal ions. The highest relative specific enzyme activity was obtained in the presence of Mn(II) ions, taken as 100% reference to all further tested ions: Mg (II) : 72%; Co(II) : 67.5%; Ni(II) : 55.9%; and Ca(II) : 47%. The enzyme showed 5.8% activity even in the presence of 20 mM EDTA, which was essentially the same as the activity of the enzyme in the absence of divalent metal ions. Zn(II) and Cu(II) ions did not activate the enzyme but on the contrary

they had slightly inhibitory effect as in the presence of Zn(II) the activity was 3.6% and in the presence of Cu(II) 3.1% (Figure 35).

Our results are similar to those described for human FUT III, human FUT V, FucT from *Silene alba,* and *C. elegans* FUT-1 (Palma AS *et al.*, 2004; Murray BW *et al.*, 1997; Léonard R *et al.*, 2005; Wilson IB, unpublished data), which are members of the GT10 family. Moreover, our results are also similar to those obtained for other fucosyltransferases, which do not belong to the GT10 family, such as *C. elegans* FUT-8, which is a core α1,6-fucosyltransferase (Paschinger K *et al.*, 2008) and core α1,6-fucosyltransferases from a number of mammalian and avian sources (Struppe and Staudacher, 2000). Core α1,6-fucosyltransferases are members of the GT23 family as well as NodZ from *H. pylori*. Its structure appears to be close to the GT-B fold (Brzezinski K *et al.*, 2007). We can conclude that divalent metal ions are not essential for the activity of fucosyltransferases. However, divalent metal ions are required to perform full enzyme activity.

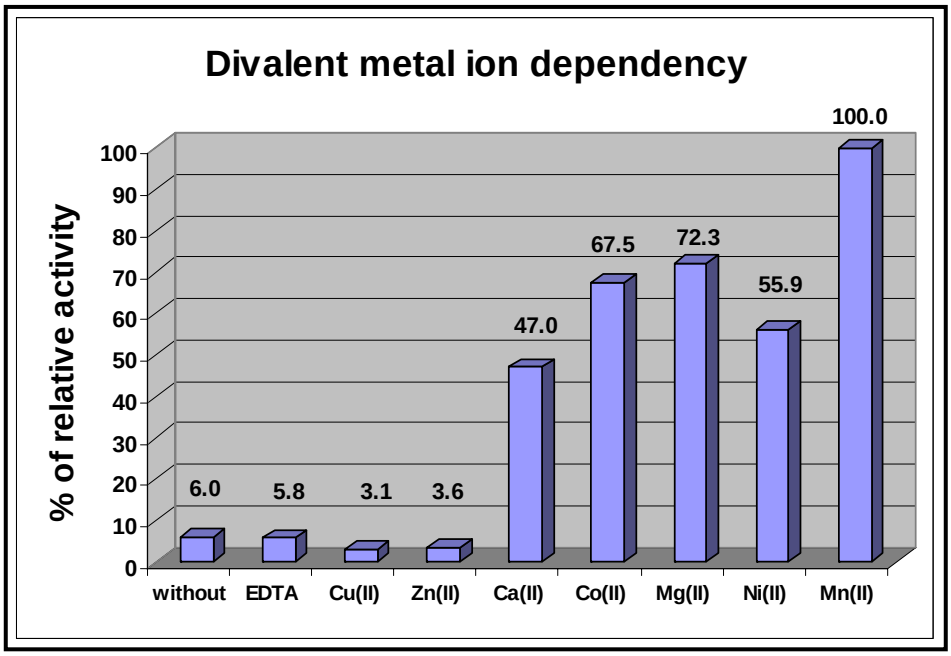


Fig. 35 **Enzyme activities of FucTA (HFΔ1-88) in the presence of divalent metal ions.** HisTrap affinity chromatography purified enzyme was further dialyzed against weak 25 mM Tris-HCl pH7.5 buffer to remove Ni(II) ions from the sample and incubated with $GnGnF^6$-NST-dansyl and GDP-Fuc either in the presence or absence of 10 mM concentration of different divalent metal ions or EDTA. Reaction conditions are described in '*Materials and Methods*' (*A. thaliana* FucTA enzyme assays). The highest enzyme activity was referred to be 100%.

3.2.3. Site-directed mutagenesis in the donor substrate binding pocket of FucTA (HFΔ1-88)

Results of our bioinformatics analysis, together with the recent crystal structure of *H. pylori* FucT, suggest that the donor substrate binding pocket of α1,3-fucosyltransferases is formed by the spatial arrangement of structural elements that correspond to the region of the catalytic domain that comprised the 1st Cluster motif and the α1,3 FucT motif, thus spanning around 80 residues.

As shown in Figures 27 and 28, the residues that make contact with the donor substrate in *Helicobacter pylori* FucT enzyme are well conserved among GT10 protein members. Since there is still little information regarding whether all conservative residues of the donor substrate binding pocket are involved in the binding of GDP-Fuc, we focused on alanine scanning mutagenesis to map the donor substrate binding domain. Specifically, we intend to elucidate importance of residues of the 1st Cluster motif and the loop region which also account for the donor substrate binding of FucTA.

As it was mention already, that there was not any solved 3D structure of a fucosyltransferase belonging to the GT10 family at the time we found the 1st Cluster motif and decided to investigate it using the tools of molecular biology. Thus our first experiments were purely based on the results of our bioinformatics analysis. Therefore, our initial experiments were aimed at the most conserved residues within this motif.

All studies were carried out on the 1st Cluster motif and the loop between the two conserved motifs of *A. thaliana* FucTA by Ala substitution of the following residues: S218, which forms hydrogen bond with the guanine, and R226, which forms hydrogen bond with the α phosphate of the GDP ligand, which both are conserved among all investigated α1,3/4-fucosyltransferases, N219, which forms a hydrogen bond with the donor substrate in our model, and Y243, which might be involved in a stacking effect between its aromatic ring and the nitrogen 7 of guanine of the donor substrate. All of these presumed interactions are deduced from crystal structures of *H. Pylori* FucT and the 3D model of *A. thaliana* FucTA (Figure 27 and 28). In the region lying between the two conserved motifs, and which seems to be a loop, alanine replacement of residue S253, which might form a hydrogen bond with the guanine of the donor substrate (Figure 36), was performed. Site-directed mutations S218A and R226A (Figure 36) were carried out on FΔ1-88 mutant, while

mutations N219A, Y243A, and S253A (Figure 36) were performed using HFΔ1-88 mutant as a template. The alanine screening often used in protein engineering projects involves the substitution of residues to alanine, based on the assumption that alanine is a `neutral' residue which do not poses steric restriction and do not change secondary structures.

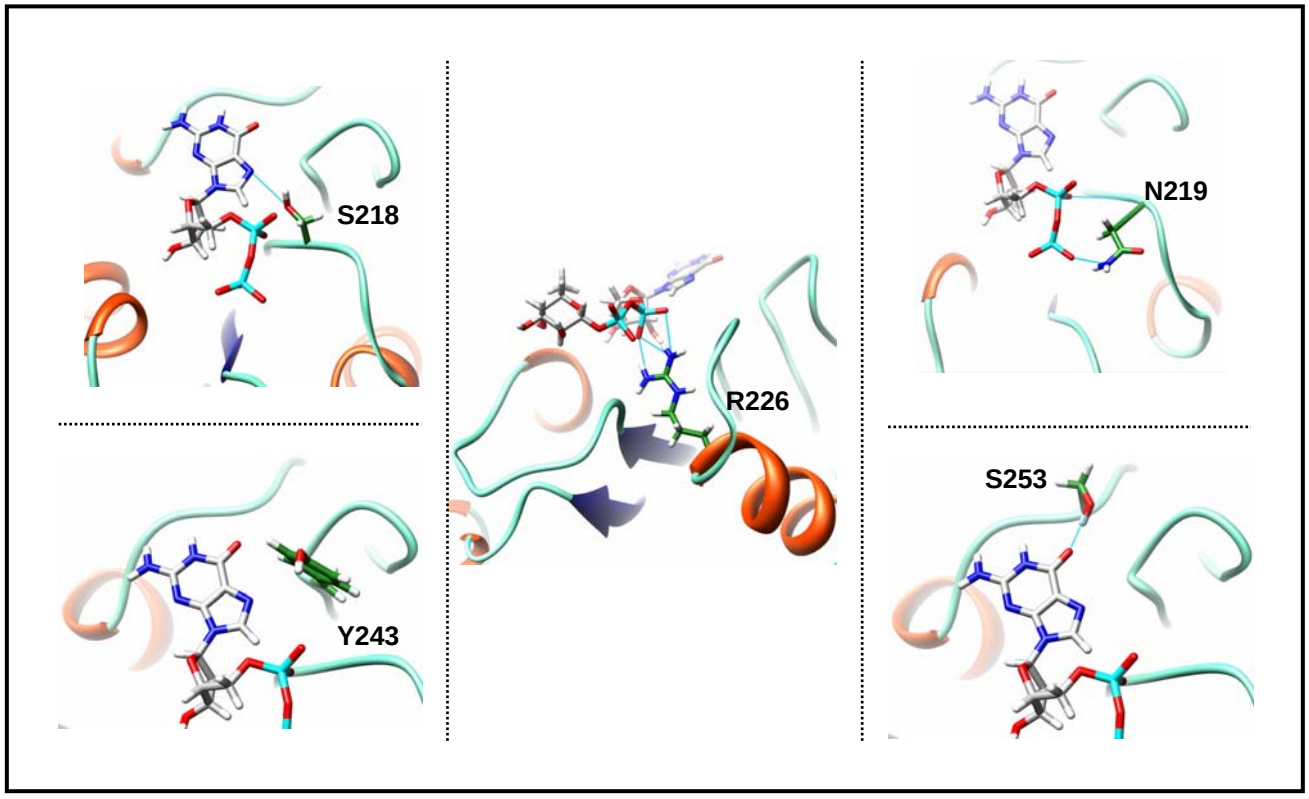


Fig. 36 **Separate view of each residue of the donor substrate binding pocket which were chosen as candidates for site-directed mutagenesis.** The pictures represent their positions and possible hydrogen bonds between their functional groups and the donor substrate (Y243 may form a stacking effect with the nitrogen 7 of the guanine).

FΔ1-88/S218A and FΔ1-88/R226A protein mutants were successfully expressed in *Pichia* with the expected molecular size (Figure 37 A). However, none of the mutants exhibited enzymatic activity (Figure 37 B). Since alanine does not impose steric effects on the protein's overall fold, probably the absence of side chain functional groups of S218 and R226 which form hydrogen bond with guanidine or phosphate groups destabilise the binding of GDP-Fuc. This is consistent with results demonstrated from others, e.g. mutation of the conserved R195 of *H.pylori* FucT which resulted in complete abolition of enzyme activity (Sun HY *et al.*, 2007).

Replacement of N219 by alanine led to a residual activity of approximately 1%. Alanine replacement of Y243 decreased the enzyme activity by approximately 90%, while mutant S253A retained around 35% of activity (Figure 37 B). Moreover, mutations Y243A and S253A do not seem to influence any secondary structure, we checked this eventuality using the Jpred 3 secondary structure prediction server (Cole C *et al.*, 2008), thus it is very probable that these residues are involved in the donor substrate binding. Of course, to prove it further experiments, such as K_m measurements, are necessary.

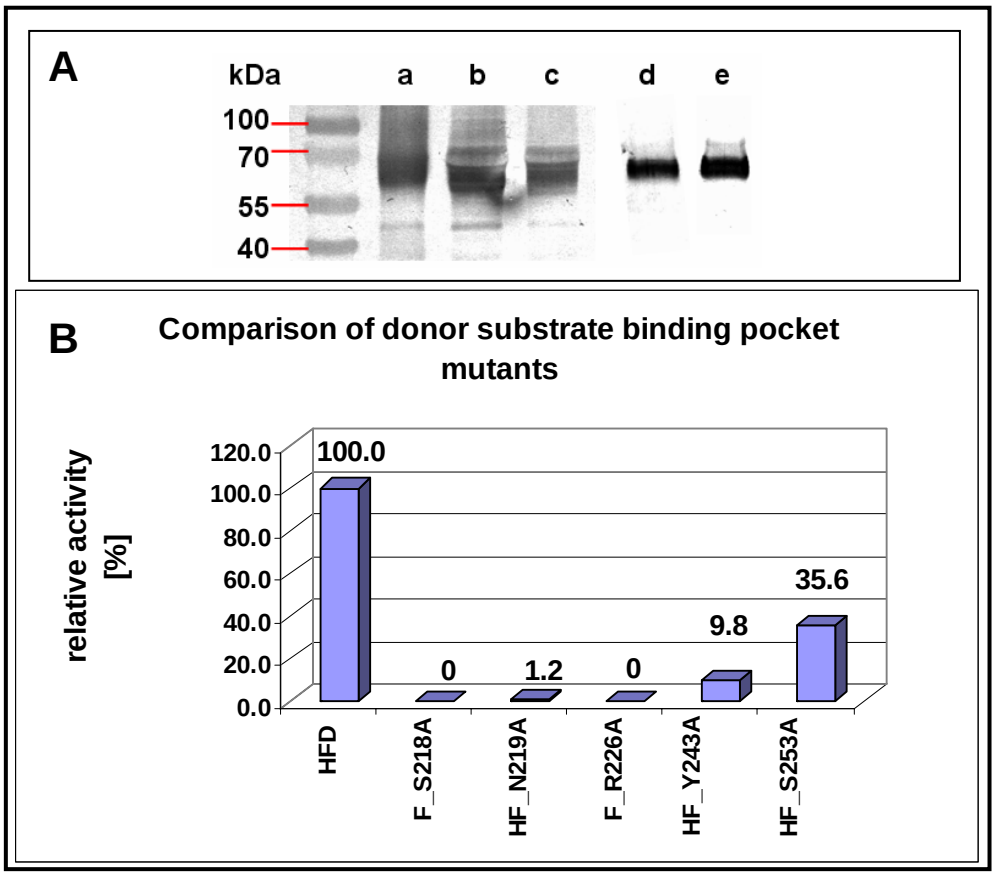


Fig. 37 **Western blot and comparison of relative specific activities of point mutated (donor substrate binding pocket) FucTAs. A: Western blotting** with mouse anti-Flag monoclonal antibodies (1:5000). FucTA proteins bearing His-tag were expressed and purified at same conditions (see Material and methods; Protein purification - HisTrap purification; Western blotting) except mutants FΔ1-88/S218A and FΔ1-88/R226A, which were methanol precipitated (300 µl supernatant + 1700 µl methanol) directly from the conditioned media. *Line a:* HFΔ1-88/N219A; *line b:* mutant HFΔ1-88/Y243A; *line c:* HFΔ1-88/S253A; *line d:* FΔ1-88/S219A; *line e:* FΔ1-88/R266A. **B: Comparison of relative specific activities of single point mutated FucTA constructs.** Activity of the HFΔ1-88 was used as reference. Enzyme activities were evaluated by reverse-phase HPLC-assisted measurement of the conversion of $GnGnF^6$-NST-dansyl (see Scheme 1). Protein concentrations (as all constructs bear Flag-tag) were estimated by anti-Flag ELISA. (*HFD*: HFΔ1-88; *F_S218A*: FΔ1-88/S218A; *HF_N219A*: HFΔ1-88/N219A; *F_R226A*: FΔ1-88/R226A; *HF_Y243A*: HFΔ1-88/Y243A; *HF_S253A*: HFΔ1-88/S253A). For conditions see *Materials and methods*, sections *A. thaliana* FucTA enzyme assays *and* Anti-Flag ELISA.

3.2.4. Putative N-glycosylation sites mutants of FucTA (HFΔ1-88)

As with many Golgi proteins, eukaryotic fucosyltransferases are N-glycosylated. It had been shown that mutations of potential N-glycosylation sites lead to decrease of enzymatic activity in the case of some GT10 members from humans (Christensen LL *et al.*, 2000 a; Christensen LL *et al.*, 2000 b).

The above-mentioned results led us to investigate the N-glycosylation of recombinant *A. thaliana* FucTA in *Pichia pastoris.*

HFΔ1-88 expressed in *P. pastoris* gives a major band of 70 kDa and another smeared band between 100 – 130 kDa (Figure 33 E), both of them showed strong anti-Flag reactivity on Western-blotting. After PNGase F treatment just traces of these bands are present while a new major band of approximately 49 kDa appears which corresponds to the theoretical Mw value of this mutant (Figure 33 E). This huge shift in the MW is due to N-glycans produced by *Pichia,* which may contain up to 50 mannosyl residues per one N-glycosylation site. In some preparations, a fainter lower band of approximately 55 kDa was also observed (Fig. 32 C); this form lacks glycosylation at the position N337 which was later confirmed by mass spectral fingerprinting (Figure 26).

HFΔ1-88 hybrid mutant bears three putative N-glycosylation sites (N337, N420, and N481). The three potential N-glycosylation sites were examined by alanine screening of the N-X-T/S motifs (Figure 30). The peptide determinant of N-glycosylation is Asn-X-Ser/Thr, where X could be any residue but proline. If either Asparagine or the Serine/Threonine is replaced by other residue the site is no longer recognised by the glycosylation machinery.

Upon expression of single N-glycosylation site (point) mutants HFΔ1-88/N337A, HFΔ1-88/T339A, HFΔ1-88/T422A, and HFΔ1-88/S483A in *P. pastoris* it became evident that N-glycosylation of one specific site is not essential for the enzyme activity. However, each mutation decreased the enzyme activity as follows: HFΔ1-88/N337A retained 18.6%, HFΔ1-88/T339A 12.1%, HFΔ1-88/T422A 32.2%, and HFΔ1-88/S483A 36.5% of the original activity (Figure 25 B).

The only N-glycosylation site mutant showing a significant shift on the western blot is HFΔ1-88/S483A. This mutant produces two very close bands around 55 kDa (Figure 38 A). Moreover, the double truncated mutant HFΔΔ1-88;388-501 expressed in *Pichia* produces two bands (very close to each other) between 40 and 55 kDa (closer to 40), while its theoretical MW is approximately 36 kDa (Figure 33 C). Since this double truncated mutant lacks the two potential C-terminal N-glycosylation

sites (N420 and N483) it suggests that the only conserved N-glycosylation site of plant core α1,3-fucosyltransferase (corresponding to N337 of *A. thaliana* FucTA) is glycosylated in this mutant.

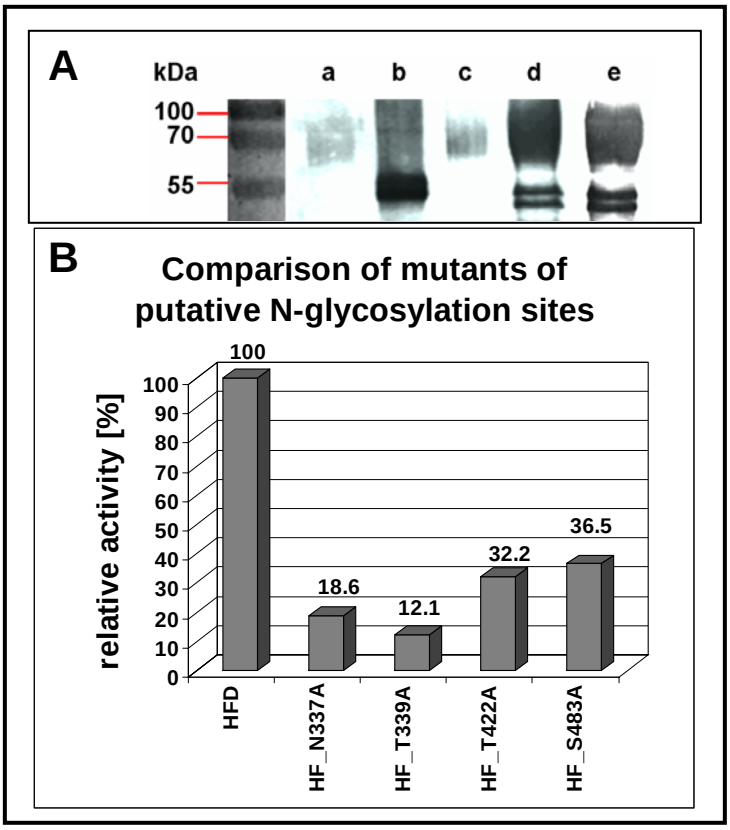


Fig. 38 **Western blot and comparison of relative specific activities of partially purified point mutated (N-glycosylation sites) FucTA. A: Western blotting** with mouse anti-Flag (1:5000). FucTA proteins were expressed and purified at same conditions (see Material and methods; HisTrap purification; Western blotting). *Line a:* HFΔ1-88; *line b:* HFΔ1-88/S483A; *line c:* HFΔ1-88/T422A; *line d:* HFΔ1-88/N337A; *line e:* HFΔ1-88/T339A. **B: Comparison of relative specific activities single point mutated FucTA (HFΔ1-88) constructs.** Activity of the HFΔ1-88 was used as a reference. Enzyme activities were evaluated by reverse-phase HPLC-assisted measurement of the conversion of $GnGnF^6$-NST-dansyl (see Scheme 1). Protein concentrations (as all constructs bear Flag-tag) were estimated by anti-Flag ELISA. (*HFD*: HFΔ1-88; *HF_N337A*: HFΔ1-88/N337A; *HF_T339A*: HFΔ1-88/T339A; *HF_T422A*: HFΔ1-88/T422A; *HF_S483A*: HFΔ1-88/S483A). For experimental procedure see *Materials and Methods* (*A. thaliana* FucTA enzyme assays *and* Anti-Flag ELISA.)

Since Western blot analysis did not answer sufficiently the question "which potential glycosylation sites are glycosylated and to which level in Pichia?", we decided to investigate the recombinant mutants and some of their glycoforms by mass spectral fingerprinting.

Up to now we have analyzed major band and the aforementioned occasionally appearing lower band as well as the PNGase F treated form of the HFΔ1-88 mutant (Figure 33 E). Furthermore, also the major band of the T339A mutant as well as the upper and lower band of the S483A mutant, both were examined (the latter are very close to each other, Figure 38 A).

Tryptic digest of the lower band of the HFΔ1-88 revealed a strong peak of 1654.8 m/z, which represents the fragment YLADNAPDAYNQTLR comprising of N337, while the upper band lacks this peak (Figure 39). This may be compatible with glycosylation of the N337 in the upper band. There is also a peak 1654.7 m/z present in the sample treated by PNGase F (Figure 39). This peak

may refer to the unglycosylated fragment as well as to its deglycosylated form YLADNAPDAYDQTLR due to PNGase F activity. A smaller peak 1643.8 m/z, which represents the fragment DGNLTLEALESAVLAK containing N420, was also identified in both bands. The 1644.84 m/z peak, which corresponds to DGDLTLEALESAVLAK is present beside the peak 1643.8 m/z in the deglycosylated sample (Figure 39; peaks are not labelled in the spectra due to label overlap), which indicates that this site is partially glycosylated.

From the spectra of HFΔ1-88 upper, lower, and PNGase F treated bands it is not evident whether the site N481 is completely or partially unoccupied (Figure 39).

In the case of the T339 mutant peak 1623.74 m/z corresponds to the mutated sequence YLADNPDAYNQALR (T339A); even this peak coincides with a peak of another fragment LRVHGIYPIGLTQR present in all samples (Figure 39). Naturally, the peak for the YLADNPDAYNQTLR fragment is missing from the spectra. Peaks representing the sites N420 and N481 are also missing; this suggests that both sites are glycosylated.

Both upper and lower bands of the S483A mutant lack the peak 1654.8 m/z, which may be compatible with glycosylation at the position N337 (Figure 40). Further, a small peak 1643.8 m/z is present in both bands, this indicates that the site N420 is not glycosylated in this case. The peak 1459.7 m/z corresponds to the mutated sequence FEGNSALSTHIQR (S483A).

P.pastoris produces incompletely glycosylated isoforms of FucTA. A similar phenomenon was observed in the case of *Vigna radiata* (mung bean) core α1,3-fucosyltransferase expressed in Sf9 insect cells. Even the enzyme possesses only two potential N-glycosylation sites, three isoforms were observed, which might correspond to un-, hemi-, and fully-glycosylated forms (Both, unpublished data)

It is evident from the results of Mass Spectral Fingerprinting and Western blot analysis that all the three N-glycosylation sites might be glycosylated but are not in all cases. It is not known whether these sites are glycosylated in *A. thaliana*. It would be interesting to prepare a triple mutant to see whether any N-glycosylation is essential for the enzyme activity (or proper folding). Furthermore, investigation of non-recombinant enzyme purified directely from *A. thaliana* would be useful.
Mass spectral fingerprinting

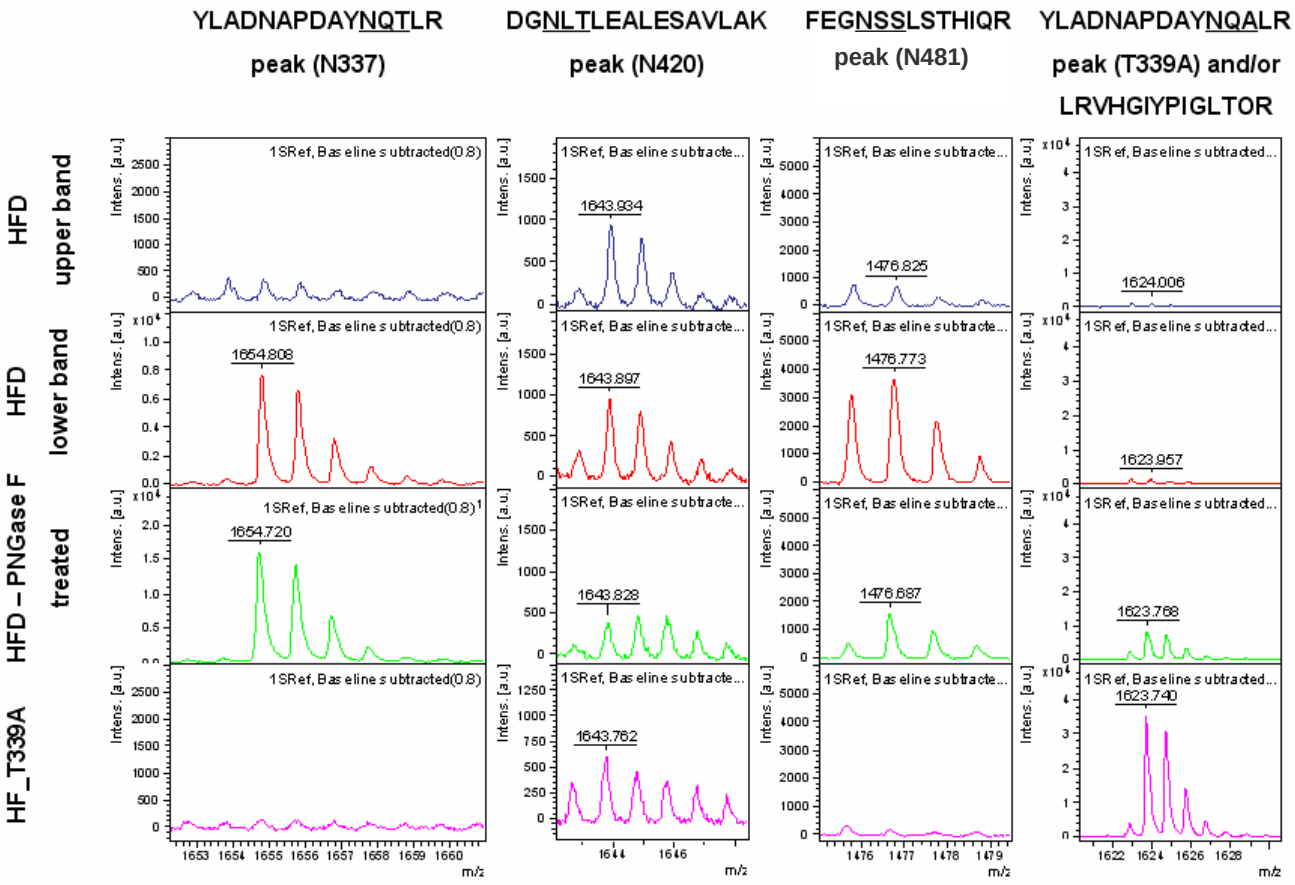


Fig. 39 **Mass spectra of Tryptic fragments representing potential N-glycosylation sites from HFΔ1-88 upper and lower band, PNGase F treated HFΔ1-88, and T339A mutant** (peak representing substitution A339 included). Cysteines were treated with iodoacetamide. It should be considered that there is an overall shift of approximately -0.1 m/z in the spectra of PNGase F treated sample.

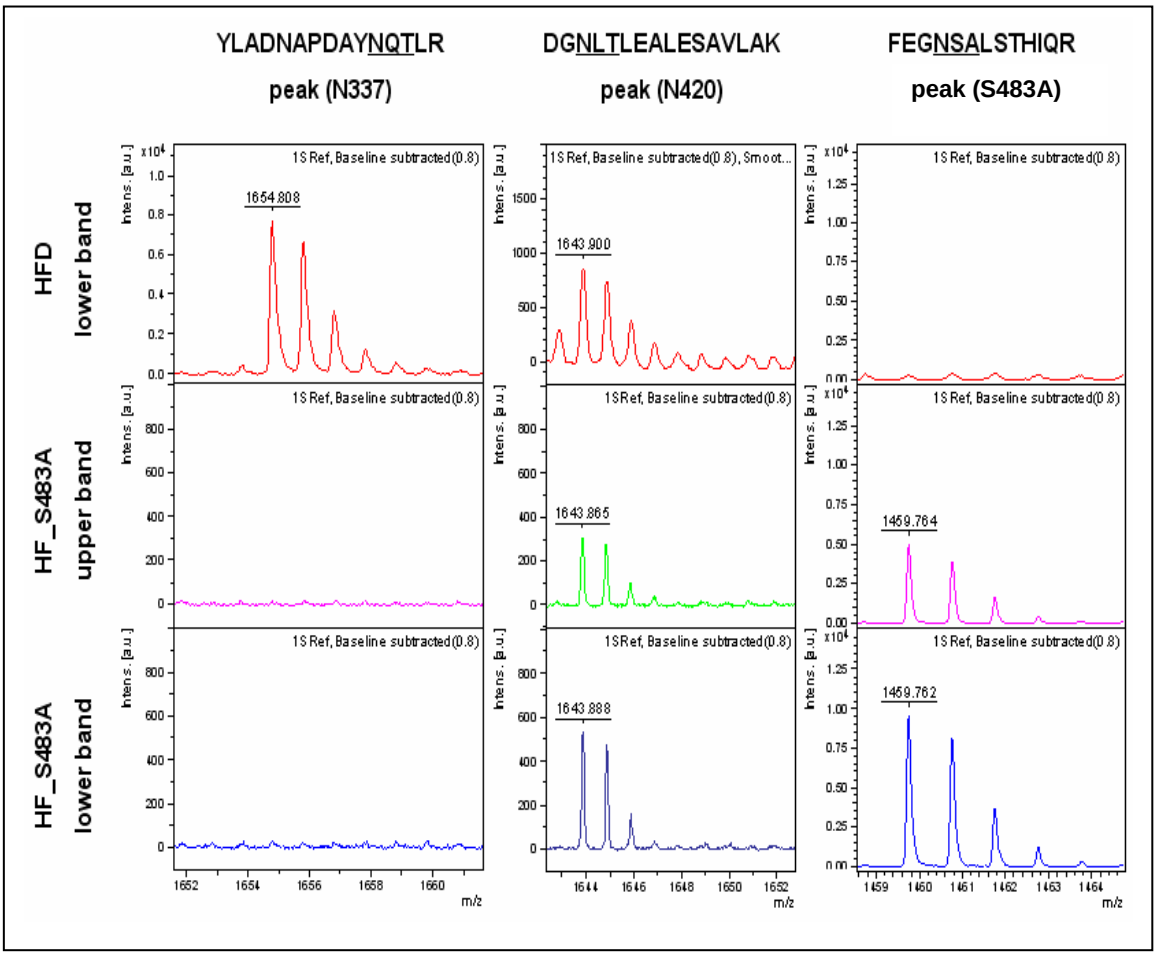


Fig. 40 **Mass spectra of Tryptic fragments representing potential N-glycosylation sites from HFΔ1-88 lower band and S483A mutant upper and lower bands** (peak representing substitution A483 included). Cysteines were treated with iodoacetamide.

4. Conclusion and Perspectives

Fucose plays an important role either as the H antigen of the H/h blood group system or in the Lewis determinants (Flowers HM, 1981). Furthermore, fucose is abundant in the photoreceptor layer of the retina of the eye (Fliesler SJ *et al.*, 1984), in the outer layer of skin (Roberts GP *and* Marks R, 1983), in the proximal tubules of the human kidney (Hennigar RA *et al.*, 1985), in the testes (Malmi R *et al.*, 1987; Domino SE *et al.*, 1989), at the junctions of nerves (Brunngraber EG *et al.*, 1975; Webster JC *and* Klingman JD, 1980), and on the surface of leukocytes and endothelium (Springer TA, 1994; Smith PL *at al.*, 1996; Lowe JB, 1997; Marquardt T *et al.*, 1999). Fucose plays an important role in the proper function of the tissues mentioned above.

Abnormal fucosylation in humans leads to diseases associated with rheumatism, type II diabetes, inflammation, neural diseases, onco-diseases (especially in the process of metastasis); also fucosylation is associated with parasitic infections, and allergies. Fucosylation events are performed by enzymes called fucosyltransferases, which catalyse the transfer of fucosyl residues from GDP-L-fucose to other carbohydrates (or proteins). In the case that the acceptor substrate is a glycoconjugate, fucosyl residue is transferred either to galactose forming an α1,2-linkage or to GlcNAc forming α1,3-, α1,4, or α1,6-linkages.

According to their sequence similarities and mode of action, fucosyltransferases responsible for the forming of α1,3/4 linkage belong to the same family (GT10). In plants as well as in some invertebrates a core α1,3-fucosylation has been described that does not occur in mammals and represents the modification of the innermost asparagine-linked GlcNAc with α1,3-linked fucose and thus forming immunogenic epitope (associated with some insect, pollen, and food allergies). However, very recently, it has been proposed that human FUT10 and FUT11 could have a similar core α1,3-fucosyltransferase activity (Mollicone R *et al.*, 2009). The predicted fold for all fucosyltransferases is GT-B or its variant. The catalytic domain of glycosyltransferases bearing this fold is built up from two subdomains, one is responsible for the binding of a specific donor substrate the second for the binding of a specific acceptor substrate.

Our aim was to investigate the protein sequences of the GT10 members (putative enzymes as well as those with known activities and biological functions) using the tools of bioinformatics with an emphasis on plant and nematode core α1,3-fucosyltransferases. We inspected sequence

similarities, conserved motifs, and potentially important residues, which may serve as potential molecular determinants for structure-function studies.

Interestingly, plant core α1,3-fucosyltransferases, compared to other GT10 members, possesses an unique additional C-terminal region of approximately 110 residues (which is highly conserved among them and shows no sequence similarity to any known protein), and our group is interested in plant core α1,3-fucosyltransferases as the first recombinant plant fucosyltransferase, which was cloned by Mucha J (Leiter H *et al.* 1999). We decided to further investigate the FucTA from *Arabidopsis thaliana*. Based on the results of bioinformatics analyses a number of structure-function studies comprising of heterologous expression, mutant preparation, and comparative analyses were proposed. We decided to examine biochemical parameters of the enzyme as well as investigate effects of protein manipulations at the N- and C-terminus of the enzyme, the role of conserved residues of the motif we have found, and the impact of glycosylation on the enzyme activity.

Recently, the crystallographic structure of FucT from *Helicobacter pylori*, a prokaryotic member of the GT10 family, obtained from crystals soaked with the donor substrate was solved (Sun HY *et al*. 2007). This fucosyltransferase, compared to *A.thaliana* FucTA, is acting on a very different acceptor substrate. However, it shows homology to *A.thaliana* FucTA (as well as to other GT10 members) in a specific region, which corresponds to the donor substrate binding pocket in the 3D structure of *H.pylori* FucT. Thus, our next step was to build a model of this region of *A.thaliana* FucTA using the tools of homology modelling and exerting the results of this structural model in structure-function studies.

Bioinformatics analyses of forty nine protein sequences of fucosyltransferases which belong to the GT10 family (retrieved from the CAZy database) showed that all investigated fucosyltransferases possess two conserved motifs. Besides the known α1,3 FucT motif we have found another motif **S-(N/H/D)-X_{5-9}-R-X_6-(L/I)-X_{3-7}-(V/L/I)-X_3-G** located very close to the previously mentioned. Since this motif lays prior to the α1,3 FucT motif we propose it as 1stCluster motif. This motif is more conserved at the level of hydrophobic clusters than at the level of primary structure that suggests rather a structural conservation. However, this motif consists of three conserved residues and some residues which have side chains of similar properties.

Furthermore, GT10 members of higher eukaryotes are also characterized by the presence of a conserved CXXC motif. Moreover, except these Cys residues there are also some other conserved Cys residues in the sequences of human, plant, and nematode α1,3-fucosyltarnsferases, which are

candidates for disulphide bridges as it was shown in the case of human α1,3-fucosyltarnsferases Fuc-TIII, V, and VII. We also identified some other conserved Cys residues, which are conserved only among invertebrate α1,3-fucosyltransferases as well as some specific to plant core α1,3-fucosyltransferases, which might be involved in disulfide bond formation. This potential disulfide bond formation could be interesting to study (in the case of *A. thaliana* FucTA), especially because of the C91 which is part of the first conserved motif among seeding plant core α1,3-fucosyltransferases. A disulfide bond between this Cys residue and for example a distant Cys residue may keep the enzyme in the correct conformation. Moreover, the results of this experiment might be applicable to other seed plant core α1,3-fucosyltransferases.

As a number of divalent metal ions play an activator role in the activity of some α1,3-fucosyltransferases we were looking for the canonic DXD motif, which is known to be responsible for the donor substrate (nucleotide-sugar) binding in glycosyltransferases adopting a GT-A fold. The two conserved residues of this motif bind the donor substrate via coordination of a divalent metal ion, which interacts with the phosphate groups of the donor. However, this inspection did not show such a conserved motif, which suggests a different mode for GDP-Fuc binding.

Multiple sequence alignment of some plant core α1,3-fucosyltransferases, which are using *GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc* as an acceptor substrate with the sequence of *C.elegans* FUT-1, which recognises *Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc* as acceptor substrate, revealed a region of approximately 20 residues prior to the conserved motifs towards the N-terminus of the investigated plant core α1,3-fucosyltransferases, which is not present in the FUT-1 sequence. This sequence might be responsible for the recognition of the distal GlcNAc in the structure of the plant acceptor substrate. Site-directed mutagenesis of residues within the mentioned region or a swapping experiment, in which this region is introduced into the sequence of FUT-1 or deleted from the sequence of FucTA, may prove our hypothesis.

Our results from structure-function studies of *A.thaliana* FucTA also demonstrate that GT10 members of higher eukaryotes are sensitive to manipulation on their C-terminus, while the N-terminal CTS domain is not essential for the enzyme activity and can be replaced by other sequences. Such a hybrid mutant of FucTA, with 88 residues deleted from its N-terminus and bearing an N-terminal 6x His+Flag-tag instead, was our subject of biochemical characterisation. The enzyme recognises either GnGn (*in vivo* and *in vitro)* or $GnGnF^6$ (*in vitro*) as acceptor substrate and initiates the forming of the HRP epitope *in vitro*. The optimal pH for the enzyme activity is pH 6.5, with optimal reaction at 30 °C. The enzyme activity is heavily stimulated by Mn^{2+} and Mg^{2+} ions,

while Cu^{2+} and Zn^{2+} ions have a slightly inhibitory effect. Using $GnGnF^6$-NST-dansyl as an acceptor substrate the enzyme's K_m value for GDP-Fuc was found to be 0.1 mM. Since the K_m value for GDP-Fuc did not change in the absence of divalent metal ions it suggests that divalent metal ions are not acting via interactions with the donor substrate. It will be useful to measure the K_m value for the acceptor substrate in the presence and absence of divalent metal ions. This is however very complicated, since the preparation of a sufficient amount of labelled acceptor substrate is very expensive and time consuming.

We were curious whether the extra C-terminal subdomain unique to plant core α1,3-fucosyltransferases is essential for FucTA activity. The deletion of this region (C388 - V501) led to a dramatic decrease of activity (<1%) but not to complete abolition. Moreover, the K_m value of this mutant for GDP-Fuc was determined to be 0.1 mM, which suggests that this part of the plant core α1,3-fucosyltransferases will rather be involved in stabilisation of the enzyme's structure and maybe in some other yet unknown functions rather than to be involved in the donor substrate binding. Since this C-terminal region is relatively long one can speculate that it may also contribute to the acceptor binding or specificity, or interaction with other Golgi proteins.

We investigated a potential function in catalysis of conserved residues (S218 and R226 of FucTA) of the 1st Cluster motif (which is delineated from S218 to G244 in the sequence of FucTA) as well as some other residues (N219, Y243) of this motif, which are predicted to interact with the donor substrate in our structural model of the donor binding pocket of FucTA. Both conserved residues S218 and R226 are essential for the enzyme activity as mutations led to a complete abolition of enzyme activity. Mutations of the other two residues led to significant decrease of enzyme activity (enzyme bearing mutation Y243A lost more then 90% and mutant N219A almost 99% of activity). Moreover, in our proposed model S253, which is part of a loop lying between the 1st Cluster motif and the α1,3 FucT motif, forms a hydrogen bond with the guanine of the donor substrate; we also investigated this residue. Enzyme bearing mutation S253A retained only 35.6 %. All realised mutations in the donor substrate binding pocket led to decrease of enzyme activity, which indicates that they are possibly involved in interactions with the donor substrate. For a more precise understanding of the effect of these mutations measurements of K_m values for the donor substrat would be very helpful.

As with many proteins processed by the Golgi apparatus, eukaryotic fucosyltransferases are N-glycosylated. It has been shown that N-glycosylation is important to maintain full activity of some human α1,3/4-fucosyltransferases. *A. thaliana* FucTA bears four putative N-glycosylation sites

(N64, N337, N420 and N481). From the heterologous expression of the above mentioned hybrid mutant FucTA in *Pichia pastoris* it was evident that the enzyme is glycosylated at least in the recombinant form. From the analysis of N-glycosylation site mutants of FucTA it is evident that N-glycosylation of one specific site is not essential for the enzyme activity. Although, mutation of any potential N-glycosylation site led to significant decrease of activity by more the 60% (except for N64, which was eliminated by truncation of the hybrid mutant that anyway had no significant effect on the enzyme activity). Mutation of the enzyme at the site N337 (which corresponds to the only N-glycosylation site conserved among plant core α1,3-fucosyltransferases) was the less active than mutation N337A, which led to a residual activity of approximately 19% and mutation T339A only 12%, while mutations of the two other potential N-glycosylation sites (N420 and N481) led to residual activities higher than 30%. Moreover, *P. pastoris* produces incompletely glycosylated isoforms of FucTA. It is evident from the results of Mass Spectral Fingerprinting and Western blot analysis that all the three N-glycosylation sites might be glycosylated but are not in all cases. To see whether any N-glycosylation is essential for the enzyme activity, prepataion of a triple N-glycosylation site mutant or a recombinant expressed in *E. coli* (which lacks N-glycosylation) is needed. As no eukaryotic α1,3-fucosyltransferase has been crystallized to date and considering the interest in human α1,3-fucosyltransferases as being necessary in the formation of selectin ligands, a crystal structure of *A. thaliana* FucTA may be a promising means to gain more information about this protein family. To this end, the preparation of an unglycosylated or hypoglycosylated, but active, enzyme would be particularly useful.

5. Abbreviations

CAZy database: Carbohydrate-Active enzymes database

CTS: N-terminal region of type II transmembrane proteins comprising cytoplasmic tail, transmembrane domain, and stem region

EDTA: ethylenediaminetetraacetic acid

ELISA: enzyme-linked immunosorbent assay

GDP-Fuc (or GDP-fucose): guanosine diphosphate-L-fucose

GlcNAc: N-acetylglucosamine

GnGn: *GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc*

HCA: Hydrophobic Cluster Analysis

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HRP: horseradish peroxidase

IgE: immunoglobulin-E (ε as heavy chain)

IgG: immunoglobulin-G (γ as heavy chain)

LB: Luria-Bertani medium

MALDI TOF: Matrix-assisted laser desorption/ionization time-of-flight

Man3 (MM): *Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc*

Mes: 2-(*N*-morpholino)ethanesulfonic acid

mQ H_2O: mQ is a brand name for deionised water by a Millipore device

MS: mass spectrometry

MWCO: molecular weight cut off

PDB: Protein Data Bank

OD: optical density

SDS: sodium dodecyl sulphate

SDS-PAGE: sodium dodecyl sulphate polyacrylamide gel electrophoresis

TFA: Trifluoroacetic acid

Tris: tris(hydroxymethyl)aminomethane

YNB: Yeast Nitrogen Base

YPD: Yeast extract Peptone Dextrose medium

YPDS: Yeast extract Peptone Dextrose Sorbitol medium

6. Literature

Acar M, **Jafar-Nejad H**, **Takeuchi H**, **Rajan A**, **Ibrani D**, **Rana NA**, **Pan H**, **Haltiwanger RS**, **Bellen HJ**; Rumi is a CAP10 domain glycosyltransferase that modifies Notch and is required for Notch signaling. Cell. **2008**;132(2):247-58.

Abeijon C, **Hirschberg CB**; Subcellular site of synthesis of the N-acetylgalactosamine (α 1-O) serine (or threonine) linkage in rat liver. J Biol Chem. **1987**; 262(9):4153-9.

Alhadeff JA; Malignant cell glycoproteins and glycolipids. Crit Rev Oncol Hematol. **1989**; 9(1):37-107.

Asao T, **Kuwano H**, **Nakamura J**, **Okamura A**, **Berger EG**, **Matta KL**, **Yazawa S**; Tumor cells as the origin of elevated serum α1,3fucosyltransferase in association with malignancy. Clin Exp Metastasis. **2000**; 18(7):605-10.

Aucoin JM, **Koul O**, **Sajdel-Sulkowska EM**, **Baboval T**, **Smith FI**; The rat α1, 3-fucosyltransferase (rFucT-IV) gene encodes both long and short forms of the enzyme which share the same intracellular location. Glycoconj J. **1998**; 15(7):671-81.

Augé C, Malleron A, Tahrat H, Marc A, Goergen JL, Cerutti M, Steelant WFA, Dellanoy P, Lubineau A; Outstanding stability of immobilised recombinant α(1-3/4)-fucosyltransferases exploited in the synthesis of Lewis a and Lewis x trisaccharides, Chem Commun. **2000**, 2017–2018.

Bakker H, **Bardor M**, **Molthoff JW**, **Gomord V**, **Elbers I**, **Stevens LH**, **Jordi W**, **Lommen A**, **Faye L**, **Lerouge P**, **Bosch D**; Galactose-extended glycans of antibodies produced by transgenic plants. Proc Natl Acad Sci U S A. **2001**; 98(5):2899-904

Bardor M, **Faveeuw C**, **Fitchette AC**, **Gilbert D**, **Galas L**, **Trottein F**, **Faye L**, **Lerouge P**; Immunoreactivity in mammals of two typical plant glyco-epitopes, core α(1,3)-fucose and core xylose. Glycobiology. **2003**; 13(6):427-34.

Bauer CH, **Reutter WG**, **Erhart KP**, **Köttgen E**, **Gerok W**; Decrease of human serum fucosyltransferase as an indicator of successful tumor therapy. Science. **1978**; 201(4362):1232-3.

Becker DJ, **Lowe JB**; Fucose: biosynthesis and biological function in mammals. Glycobiology. **2003**; 13(7):41R-53R.

Bencúrová M, **Rendić D**, **Fabini G**, **Kopecky EM**, **Altmann F**, **Wilson IBH**; Expression of eukaryotic glycosyltransferases in the yeast Pichia pastoris. Biochimie. **2003**; 85(3-4):413-22.

Bengtson P, **Larson C**, **Lundblad A**, **Larson G**, **Påhlsson P**; Identification of a missense mutation (G329A;Arg(110)--> Gln) in the human FUT7 gene. J Biol Chem. **2001**; 276(34):31575-82.

Breton C, **Oriol R**, **Imberty A**; Conserved structural features in eukaryotic and prokaryotic fucosyltransferases. Glycobiology. **1998**; 8(1):87-94.

Breton C, **Snajdrová L**, **Jeanneau C**, **Koca J**, **Imberty A**; Structures and mechanisms of glycosyltransferases. Glycobiology. **2006**; 16(2):29R-37R.

Brunngraber EG, **Brown BD**, **Aro A**; Distribution and age-dependent concentration in brain tissue of glycoproteins containing N-acetylgalactosamine. Neurobiology. **1975**; 5(6):339-46.

Brückner K, **Perez L**, **Clausen H**, **Cohen S**; Glycosyltransferase activity of Fringe modulates Notch-Delta interactions. Nature. **2000**; 406(6794):411-5.

Brzezinski K, **Stepkowski T**, **Panjikar S**, **Bujacz G**, **Jaskolski M**; High-resolution structure of NodZ fucosyltransferase involved in the biosynthesis of the nodulation factor. Acta Biochim Pol. **2007**; 54(3):537-49.

Burda P, **Aebi M**; The dolichol pathway of N-linked glycosylation. Biochim Biophys Acta. **1999**; 1426(2):239-57.

Callebaut I, **Labesse G**, **Durand P**, **Poupon A**, **Canard L**, **Chomilier J**, **Henrissat B**, **Mornon JP**; Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell Mol Life Sci. **1997**; 53(8):621-45.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B; The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. **2009**; 37(Database issue):D233-8.

Chazalet V, **Uehara K**, **Geremia RA**, **Breton C**; Identification of essential amino acids in the Azorhizobium caulinodans fucosyltransferase NodZ. J Bacteriol. **2001**; 183(24):7067-75.

Chen YJ, **Wing DR**, **Guile GR**, **Dwek RA**, **Harvey DJ**, **Zamze S**; Neutral N-glycans in adult rat brain tissue--complete characterisation reveals fucosylated hybrid and complex structures. Eur J Biochem. **1998**; 251(3):691-703.

Christensen LL, **Jensen UB**, **Bross P**, **Ørntoft TF**; The C-terminal N-glycosylation sites of the human α1,3/4-fucosyltransferase III, -V, and -VI (hFucTIII, -V, and -VI) are necessary for the expression of full enzyme activity. Glycobiology. **2000**; 10(9):931-9.

Christensen LL, **Bross P**, **Ørntoft TF**; Glycosylation of the N-terminal potential N-glycosylation sites in the human α1,3-fucosyltransferase V and -VI (hFucTV and -VI). Glycoconj J. **2000**; 17(12):859-65.

Cieniewski-Bernard C, **Mounier Y**, **Michalski JC**, **Bastide B**; O-GlcNAc level variations are associated with the development of skeletal muscle atrophy. J Appl Physiol. **2006**; 100(5):1499-505.

Cole C, **Barber JD**, **Barton GJ**; The Jpred 3 secondary structure prediction server. Nucleic Acids Res. **2008**; 36(Web Server issue):W197-201.

Collins BE, **Paulson JC**; Cell surface biology mediated by low affinity multivalent protein-glycan interactions. Curr Opin Chem Biol. **2004**; 8(6):617-25.

Coutinho PM, **Henrissat B**; Life with no sugars? J Mol Microbiol Biotechnol. **1999**; 1(2):307-8.

Cregg JM, **Vedvick TS**, **Raschke WC**; Recent advances in the expression of foreign genes in Pichia pastoris. Biotechnology (NY). **1993**; 11(8):905-10.

Cudney B; Protein crystallization and dumb luck. Rigaku Journal. **1999**; 16(1) :1-7.

Domino SE, **Bocckino SB**, **Garbers DL**; Activation of phospholipase D by the fucose-sulfate glycoconjugate that induces an acrosome reaction in spermatozoa. J Biol Chem. **1989**; 264(16):9412-9.

Dupuy F, **Petit JM**, **Mollicone R**, **Oriol R**, **Julien R**, **Maftah A**; A single amino acid in the hypervariable stem domain of vertebrate α1,3/1,4-fucosyltransferases determines the type 1/type 2 transfer. Characterization of acceptor substrate specificity of the lewis enzyme by site-directed mutagenesis. J Biol Chem. **1999**; 274(18):12257-62.

Dupuy F, **Germot A**, **Marenda M**, **Oriol R**, **Blancher A**, **Julien R**, **Maftah A**; α1,4-fucosyltransferase activity: a significant function in the primate lineage has appeared twice independently. Mol Biol Evol. **2002**;19(6):815-24.

Dupuy F, **Germot A**, **Julien R**, **Maftah A**; Structure/function study of Lewis α3- and α3/4-fucosyltransferases: the α1,4 fucosylation requires an aromatic residue in the acceptor-binding domain. Glycobiology. **2004**;14(4):347-56.

Endo T, **Toda T**; Glycosylation in congenital muscular dystrophies. Biol Pharm Bull. **2003**; 26(12):1641-7.

Ernst JF, **Prill SK**; O-glycosylation. Med Mycol. **2001**; 39 Suppl 1:67-74.

Eswar N, **Webb B**, **Marti-Renom MA**, **Madhusudhan MS**, **Eramian D**, **Shen MY**, **Pieper U**, **Sali A**; Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci. **2007** Nov; Chapter 2:Unit 2.9.

Fabini G, **Freilinger A**, **Altmann F**, **Wilson IB**; Identification of core α1,3-fucosylated glycans and cloning of the requisite fucosyltransferase cDNA from Drosophila melanogaster. Potential basis of the neural anti-horseadish peroxidase epitope. J Biol Chem. **2001**; 276(30):28058-67.

Faveeuw C, **Mallevaey T**, **Paschinger K**, **Wilson IB**, **Fontaine J**, **Mollicone R**, **Oriol R**, **Altmann F**, **Lerouge P**, **Capron M**, **Trottein F**; Schistosome N-glycans containing core α3-fucose and core beta 2-xylose epitopes are strong inducers of Th2 responses in mice. Eur J Immunol. **2003**; 33(5):1271-81.

Fliesler SJ, **Tabor GA**, **Hollyfield JG**; Glycoprotein synthesis in the human retina: localization of the lipid intermediate pathway. Exp Eye Res. **1984**; 39(2):153-73.

Flowers HM; Chemistry and biochemistry of D- and L-fucose. Adv Carbohydr Chem Biochem. **1981**; 39:279-345.

Flögel M, **Lauc G**, **Gornik I**, **Macek B**; Fucosylation and galactosylation of IgG heavy chains differ between acute and remission phases of juvenile chronic arthritis. Clin Chem Lab Med. **1998**; 36(2):99-102.

Furukawa K, **Matsuta K**, **Takeuchi F**, **Kosuge E**, **Miyamoto T**, **Kobata A**; Kinetic study of a galactosyltransferase in the B cells of patients with rheumatoid arthritis. Int Immunol. **1990**; 2(1):105-12.

Furukawa K, **Takamiya K**, **Furukawa K**; β1,4-N-acetylgalactosaminyltransferase--GM2/GD2 synthase: a key enzyme to control the synthesis of brain-enriched complex gangliosides. Biochim Biophys Acta. **2002**; 1573(3):356-62.

Gaboriaud C, **Bissery V**, **Benchetrit T**, **Mornon JP**; Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. FEBS Lett. **1987**; 224(1):149-55.

Galesi AL, **Aguiar MA**, **Astray RM**, **Augusto EF**, **Moraes AM**; Growth of recombinant Drosophila melanogaster Schneider 2 cells producing rabies virus glycoprotein in bioreactor employing serum-free medium. Cytotechnology. **2008**; 57(1):73-81.

Gluzman Y; SV40-transformed simian cells support the replication of early SV40 mutants. Cell. **1981**; 23(1):175-82.

Goddard TD, **Huang CC**, **Ferrin TE**; Software extensions to UCSF chimera for interactive visualization of large molecular assemblies. Structure. **2005**; 13(3):473-82

Gornik I, **Maravić G**, **Dumić J**, **Flögel M**, **Lauc G**; Fucosylation of IgG heavy chains is increased in rheumatoid arthritis. Clin Biochem. **1999**; 32(8):605-8.

Grabenhorst E, **Conradt HS**; The cytoplasmic, transmembrane, and stem regions of glycosyltransferases specify their in vivo functional sublocalization and stability in the Golgi. J Biol Chem. **1999**; 274(51):36107-16.

Hakomori S; Aberrant glycosylation in tumors and tumor-associated carbohydrate antigens. Adv Cancer Res. **1989**; 52:257-331.

Hanisch FG; O-glycosylation of the mucin type. Biol Chem. **2001**; 382(2):143-9.

Hart GW, **Housley MP**, **Slawson C**; Cycling of O-linked β-N-acetylglucosamine on nucleocytoplasmic proteins. Nature. **2007**; 446(7139):1017-22.

Helenius A, **Aebi M**; Roles of N-linked glycans in the endoplasmic reticulum. Annu Rev Biochem. **2004**; 73:1019-49.

Hemmer W, **Focke M**, **Kolarich D**, **Wilson IB**, **Altmann F**, **Wöhrl S**, **Götz M**, **Jarisch R**; Antibody binding to venom carbohydrates is a frequent cause for double positivity to honeybee and yellow jacket venom in patients with stinging-insect allergy. J Allergy Clin Immunol. **2001**; 108(6):1045-52.

Hennigar RA, **Schulte BA**, **Spicer SS**; Heterogeneous distribution of glycoconjugates in human kidney tubules. Anat Rec. **1985**; 211(4):376-90.

Hitchman RB, Possee RD, King LA; Baculovirus expression systems for recombinant protein production in insect cells. Recent Pat Biotechnol. **2009**; 3(1):46-54.

Hoja-Lukowicz D, **Ciołczyk D**, **Bergquist J**, **Lityńska A**, **Laidler P**; High-mannose-type oligosaccharides from human placental arylsulfatase A are core fucosylated as confirmed by MALDI MS. Glycobiology. **2000**; 10(6):551-7.

Hokke CH, **Yazdanbakhsh M**; Schistosome glycans and innate immunity. Parasite Immunol. **2005**; 27(7-8):257-64.

Holmes EH, **Xu Z**, **Sherwood AL**, **Macher BA**; Structure-function analysis of human $\alpha 1 \rightarrow 3$fucosyltransferases. A GDP-fucose-protected, N-ethylmaleimide-sensitive site in FucT-III and FucT-V corresponds to Ser178 in FucT-IV. J Biol Chem. **1995**; 270(14):8145-51.

Holmes EH, **Yen TY**, **Thomas S**, **Joshi R**, **Nguyen A**, **Long T**, **Gallet F**, **Maftah A**, **Julien R**, **Macher BA**; Human α1,3/4 fucosyltransferases. Characterization of highly conserved cysteine residues and N-linked glycosylation sites. J Biol Chem. **2000**; 275(32):24237-45.

Hooper LV, **Gordon JI**; Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity. Glycobiology. **2001**;11(2):1R-10R.

Hu Y, **Walker S**; Remarkable structural similarities between diverse glycosyltransferases. Chem Biol. **2002**; 9(12):1287-96.

Ihara H, **Ikeda Y**, **Toma S**, **Wang X**, **Suzuki T**, **Gu J**, **Miyoshi E**, **Tsukihara T**, **Honke K**, **Matsumoto A**, **Nakagawa A**, **Taniguchi N**; Crystal structure of mammalian α1,6-fucosyltransferase, FUT8. Glycobiology. **2007**; 17(5):455-66.

Ioffe E, Liu Y, Stanley P; Complex N-glycans in Mgat1 null preimplantation embryos arise from maternal Mgat1 RNA. Glycobiology. **1997**; 7(7):913-9.

Jacobs PP, **Callewaert N**; N-glycosylation engineering of biopharmaceutical expression systems. Curr Mol Med. **2009**; 9(7):774-800

JENSEN FC, GIRARDI AJ, GILDEN RV, KOPROWSKI H; INFECTION OF HUMAN AND SIMIAN TISSUE CULTURES WITH ROUS SARCOMA VIRUS. Proc Natl Acad Sci U S A. **1964**; 52:53-9.

Jost F, **de Vries T**, **Knegtel RM**, **Macher BA**; Mutation of amino acids in the α1,3-fucosyltransferase motif affects enzyme activity and Km for donor and acceptor substrates. Glycobiology. **2005**; 15(2):165-75.

Keegstra K, **Raikhel N**; Plant glycosyltransferases. Curr Opin Plant Biol. **2001**; 4(3):219-24.

Kelly RJ, **Rouquier S**, **Giorgi D**, **Lennon GG**, **Lowe JB**; Sequence and expression of a candidate for the human Secretor blood group α(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. J Biol Chem. **1995**; 270(9):4640-9.

Kijima-Suda I, **Miyazawa T**, **Itoh M**, **Toyoshima S**, **Osawa T**; Possible mechanism of inhibition of experimental pulmonary metastasis of mouse colon adenocarcinoma 26 sublines by a sialic acid: nucleoside conjugate. Cancer Res. **1988**; 48(13):3728-32.

Kim JH, **Lingwood CA**, **Williams DB**, **Furuya W**, **Manolson MF**, **Grinstein S**; Dynamic measurement of the pH of the Golgi complex in living cells using retrograde transport of the verotoxin receptor. J Cell Biol. **1996**; 134(6):1387-99.

Kingsley PD, **Hagen KG**, **Maltby KM**, **Zara J**, **Tabak LA**; Diverse spatial expression patterns of UDP-GalNAc:polypeptide N-acetylgalactosaminyl-transferase family member mRNAs during mouse development. Glycobiology. **2000**; 10(12):1317-23.

Konrad RJ, **Kudlow JE**; The role of O-linked protein glycosylation in beta-cell dysfunction. Int J Mol Med. **2002**; 10(5):535-9.

Kumar S, **Nei M**, **Dudley J**, **Tamura K**; MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. **2008**; 9(4):299-306.

Lairson LL, **Henrissat B**, **Davies GJ**, **Withers SG**; Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem. **2008**; 77:521-55.

Larsen RD, **Ernst LK**, **Nair RP**, **Lowe JB**; Molecular cloning, sequence, and expression of a human GDP-L-fucose:β-D-galactoside 2-α-L-fucosyltransferase cDNA that can form the H blood group antigen. Proc Natl Acad Sci U S A. **1990**; 87(17):6674-8.

Lee SY; High cell-density culture of Escherichia coli. Trends Biotechnol. **1996**; 14(3):98-105.

Legault DJ, **Kelly RJ**, **Natsuka Y**, **Lowe JB**; Human α(1,3/1,4)-fucosyltransferases discriminate between different oligosaccharide acceptor substrates through a discrete peptide fragment. J Biol Chem. **1995**; 270(36):20987-96.

Leiter H, **Mucha J**, **Staudacher E**, **Grimm R**, **Glössl J**, **Altmann F**; Purification, cDNA cloning, and expression of GDP-L-Fuc:Asn-linked GlcNAc α1,3-fucosyltransferase from mung beans. J Biol Chem. **1999**; 274(31):21830-9.

Léonard R, **Lhernould S**, **Carlué M**, **Fleurat P**, **Maftah A**, **Costa G**; Biochemical characterization of Silene alba α4-fucosyltransferase and Lewis a products. Glycoconj J. **2005**; 22(1-2):71-8.

Lerouge P, **Cabanes-Macheteau M**, **Rayon C**, **Fischette-Lainé AC**, **Gomord V**, **Faye L**; N-glycoprotein biosynthesis in plants: recent developments and future trends. Plant Mol Biol. **1998**; 38(1-2):31-48.

Lerouge P, **Bardor M**, **Pagny S**, **Gomord V**, **Faye L**; N-glycosylation of recombinant pharmaceutical glycoproteins produced in transgenic plants: towards an humanisation of plant N-glycans. Curr Pharm Biotechnol. **2000**; 1(4):347-54.

Li P, Anumanthan A, Gao XG, Ilangovan K, Suzara VV, Düzgüneş N, Renugopalakrishnan V; Expression of recombinant proteins in Pichia pastoris. Appl Biochem Biotechnol. **2007**; 142(2):105-24.

Liu F, **Iqbal K**, **Grundke-Iqbal I**, **Hart GW**, **Gong CX**; O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. Proc Natl Acad Sci U S A. **2004**;101(29):10804-9.

Llopis J, **McCaffery JM**, **Miyawaki A**, **Farquhar MG**, **Tsien RY**; Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. Proc Natl Acad Sci USA. **1998**; 95(12):6803-8.

Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky L, Darnell J; Molecular Cell Biology, Fifth Edition. **2004**, Freeman books

Lommel M, Strahl S; Protein O-mannosylation: conserved from bacteria to humans. Glycobiology. **2009**;19(8):816-28.

Lowe JB; The blood group-specific human glycosyltransferases. Baillieres Clin Haematol. **1993**; 6(2):465-92.

Macher BA, **Holmes EH**, **Swiedler SJ**, **Stults CL**, **Srnka CA**; Human α1-3 fucosyltransferases. Glycobiology. **1991**; 1(6):577-84.

Majewska-Sawka A, **Nothnagel EA**; The multiple roles of arabinogalactan proteins in plant development. Plant Physiol. **2000**; 122(1):3-10.

Malmi R, **Kallajoki M**, **Suominen J**; Distribution of glycoconjugates in human testis. A histochemical study using fluorescein- and rhodamine-conjugated lectins. Andrologia. **1987**; 19(3):322-32.

Mann PR, **Williams RH**, **Gray GM**; Distribution of glycoproteins containing fucose in normal and psoriatic keratinocytes. Br J Dermatol. **1980**; 102(6):649-57.

Marquardt T, **Brune T**, **Lühn K**, **Zimmer KP**, **Körner C**, **Fabritz L**, **van der Werft N**, **Vormoor J**, **Freeze HH**, **Louwen F**, **Biermann B**, **Harms E**, **von Figura K**, **Vestweber D**, **Koch HG**; Leukocyte adhesion deficiency II syndrome, a generalized defect in fucose metabolism. J Pediatr. **1999**; 134(6):681-8.

Martin SL, **Edbrooke MR**, **Hodgman TC**, **van den Eijnden DH**, **Bird MI**; Lewis X biosynthesis in Helicobacter pylori. Molecular cloning of an α(1,3)-fucosyltransferase gene. J Biol Chem. **1997**; 272(34):21349-56.

Martinez-Duncker I, **Mollicone R**, **Candelier JJ**, **Breton C**, **Oriol R**; A new superfamily of protein-O-fucosyltransferases, α2-fucosyltransferases, and α6-fucosyltransferases: phylogeny and identification of conserved peptide motifs. Glycobiology. **2003**; 13(12):1C-5C.

Martinez-Fleites C, **Macauley MS**, **He Y**, **Shen DL**, **Vocadlo DJ**, **Davies GJ**; Structure of an O-GlcNAc transferase homolog provides insight into intracellular glycosylation. Nat Struct Mol Biol. **2008**; 15(7):764-5.

Mason CM, **Azizi SQ**, **Dal Nogare AR**; Respiratory epithelial carbohydrate levels of rats with gram-negative bacillary colonization. J Lab Clin Med. **1992**; 120(5):740-5.

Molinari M, **Galli C**, **Vanoni O**, **Arnold SM**, **Kaufman RJ**; Persistent glycoprotein misfolding activates the glucosidase II/UGT1-driven calnexin cycle to delay aggregation and loss of folding competence. Mol Cell. **2005**; 20(4):503-12.

Mollicone R, **Reguigne I**, **Kelly RJ**, **Fletcher A**, **Watt J**, **Chatfield S**, **Aziz A**, **Cameron HS**, **Weston BW**, **Lowe JB**; Molecular basis for Lewis α(1,3/1,4)-fucosyltransferase gene deficiency (FUT3) found in Lewis-negative Indonesian pedigrees. J Biol Chem. **1994**; 269(33):20987-94.

Moloney DJ, Panin VM, Johnston SH, Chen J, Shao L, Wilson R, Wang Y, Stanley P, Irvine KD, Haltiwanger RS, Vogt TF; Fringe is a glycosyltransferase that modifies Notch. Nature. **2000**; 406(6794):369-75.

Moran AP, **Knirel YA**, **Senchenkova SN**, **Widmalm G**, **Hynes SO**, **Jansson PE**; Phenotypic variation in molecular mimicry between Helicobacter pylori lipopolysaccharides and human gastric epithelial cell surface glycoforms. Acid-induced phase variation in Lewis(x) and Lewis(y) expression by H. Pylori lipopolysaccharides. J Biol Chem. **2002**; 277(8):5785-95.

Moréra S, **Larivière L**, **Kurzeck J**, **Aschke-Sonnenborn U**, **Freemont PS**, **Janin J**, **Rüger W**; High resolution crystal structures of T4 phage β-glucosyltransferase: induced fit and effect of substrate and metal binding. J Mol Biol. **2001**; 311(3):569-77.

Mollicone R, **Moore SE**, **Bovin N**, **Garcia-Rosasco M**, **Candelier JJ**, **Martinez-Duncker I**, **Oriol R**; Activity, splice variants, conserved peptide motifs, and phylogeny of two new alpha1,3-fucosyltransferase families (FUT10 and FUT11). J Biol Chem. **2009**; 284(7):4723-38.

Murray BW, **Wittmann V**, **Burkart MD**, **Hung SC**, **Wong CH**; Mechanism of human α-1,3-fucosyltransferase V: glycosidic cleavage occurs prior to nucleophilic attack. Biochemistry. **1997**; 36(4):823-31.

Murrey HE, **Gama CI**, **Kalovidouris SA**, **Luo WI**, **Driggers EM**, **Porton B**, **Hsieh-Wilson LC**; Protein fucosylation regulates synapsin Ia/Ib expression and neuronal morphology in primary hippocampal neurons. Proc Natl Acad Sci USA. **2006**; 103(1):21-6.

Okajima T, **Xu A**, **Irvine KD**; Modulation of notch-ligand binding by protein O-fucosyltransferase 1 and fringe. J Biol Chem. **2003**; 278(43):42340-5.

Oriol R, **Mollicone R**, **Cailleau A**, **Balanzino L**, **Breton C**; Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria. Glycobiology. **1999**; 9(4):323-34.

Packer NH, **von der Lieth CW**, **Aoki-Kinoshita KF**, **Lebrilla CB**, **Paulson JC**, **Raman R**, **Rudd P**, **Sasisekharan R**, **Taniguchi N**, **York WS**; Frontiers in glycomics: bioinformatics and biomarkers in disease. An NIH white paper prepared from discussions by the focus groups at a workshop on the NIH campus, Bethesda MD (September 11-13, 2006). Proteomics. **2008**; 8(1):8-20.

Pak JE, **Rini JM**; X-ray crystal structure determination of mammalian glycosyltransferases. Methods Enzymol. **2006**; 416:30-48.

Palma AS, **Vila-Verde C**, **Pires AS**, **Fevereiro PS**, **Costa J**; A novel plant α4-fucosyltransferase (Vaccinium myrtillus L.) synthesises the Lewis(a) adhesion determinant. FEBS Lett. **2001**; 499(3):235-8.

Palma AS, **Morais VA**, **Coelho AV**, **Costa J**; Effect of the manganese ion on human α3/4 fucosyltransferase III activity. Biometals. **2004**; 17(1):35-43.

Panin VM, **Shao L**, **Lei L**, **Moloney DJ**, **Irvine KD**, **Haltiwanger RS**; Notch ligands are substrates for protein O-fucosyltransferase-1 and Fringe. J Biol Chem. **2002**; 277(33):29945-52.

Paschinger K, **Rendic D**, **Lochnit G**, **Jantsch V**, **Wilson IBH**; Molecular basis of anti-horseradish peroxidase staining in Caenorhabditis elegans. J Biol Chem. **2004**; 279(48):49588-98.

Paschinger K, **Staudacher E**, **Stemmer U**, **Fabini G**, **Wilson IBH**; Fucosyltransferase substrate specificity and the order of fucosylation in invertebrates. Glycobiology. **2005**; 15(5):463-74.

Paschinger K, **Gutternigg M**, **Rendić D**, **Wilson IBH**; The N-glycosylation pattern of Caenorhabditis elegans. Carbohydr Res. **2008**; 343(12):2041-9.

Peterson RE, **McClay DR** ; A Fringe-modified Notch signal affects specification of mesoderm and endoderm in the sea urchin embryo. Dev Biol. **2005**; 282(1):126-37.

Pettersen EF, **Goddard TD**, **Huang CC**, **Couch GS**, **Greenblatt DM**, **Meng EC**, **Ferrin TE**; UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. **2004**; 25(13):1605-12

Rampal R, **Li AS**, **Moloney DJ**, **Georgiou SA**, **Luther KB**, **Nita-Lazar A**, **Haltiwanger RS**; Lunatic fringe, manic fringe, and radical fringe recognize similar specificity determinants in O-fucosylated epidermal growth factor-like repeats. J Biol Chem. **2005**; 280(51):42454-63.

Rechsteiner M, **Rogers SW**; PEST sequences and regulation by proteolysis. Trends Biochem Sci. **1996**; 21(7):267-71.

Reid ME, Lomas-Francis C; The Blood Group Antigen Facts Book. Second ed. **2004**, New York: Elsevier Academic Press

Roberts GP, **Marks R**; The distribution of precursors of glycoconjugate synthesis in human epidermis. Arch Dermatol Res. **1983**; 275(3):199-201.

Saint-Jore-Dupas C, **Faye L**, **Gomord V** ; From planta to pharma with glycosylation in the toolbox. Trends Biotechnol. **2007**; 25(7):317-23.

Sander L, Harrysson A; Using cell size kinetics to determine optimal harvest time for Spodoptera frugiperda and Trichoplusia ni BTI-TN-5B1-4 cells infected with a baculovirus expression vector system expressing enhanced green fluorescent protein. Cytotechnology. **2007**; 54(1):35-48.

Sanner MF, **Olson AJ**, **Spehner JC**; Reduced surface: an efficient way to compute molecular surfaces. Biopolymers. **1996**; 38(3):305-20.

Seksek O, **Biwersi J**, **Verkman AS**; Direct measurement of trans-Golgi pH in living cells and regulation by second messengers. J Biol Chem. **1995**; 270(10):4967-70.

Serafini-Cessi F, **Malagolini N**, **Dall'Olio F**; A tetraantennary glycopeptide from human Tamm-Horsfall glycoprotein inhibits agglutination of desialylated erythrocytes induced by leucoagglutinin. Biosci Rep. **1984**; 4(11):973-8.

Shao L, **Luo Y**, **Moloney DJ**, **Haltiwanger R**; O-glycosylation of EGF repeats: identification and initial characterization of a UDP-glucose: protein O-glucosyltransferase. Glycobiology. **2002**; 12(11):763-70.

Showalter AM; Structure and function of plant cell wall proteins. Plant Cell. **1993**; 5(1):9-23.

Shao L, **Haltiwanger RS**; O-fucose modifications of epidermal growth factor-like repeats and thrombospondin type 1 repeats: unusual modifications in unusual places. Cell Mol Life Sci. **2003**; 60(2):241-50.

Shapiro AL, **Viñuela E**, **Maizel JV Jr**; Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. Biochem Biophys Res Commun. **1967**; 28(5):815-20.

Sherwood AL, **Nguyen AT**, **Whitaker JM**, **Macher BA**, **Stroud MR**, **Holmes EH** ; Human α1,3/4-fucosyltransferases. III. A Lys/Arg residue located within the alpha1,3-FucT motif is required for activity but not substrate binding. J Biol Chem. **1998**; 273(39):25256-60.

Sherwood AL, **Holmes EH**; Analysis of the expression and enzymatic properties of α1→3fucosyltransferase from human lung carcinoma NCI-H69 and PC9 cells. Glycobiology. **1999**; 9(7):637-43.

Shinoda K, **Morishita Y**, **Sasaki K**, **Matsuda Y**, **Takahashi I**, **Nishi T**; Enzymatic characterization of human α1,3-fucosyltransferase Fuc-TVII synthesized in a B cell lymphoma cell line. J Biol Chem. **1997**; 272(51):31992-7.

Sourrouille C, **Marquet-Blouin E**, **D'Aoust MA**, **Kiefer-Meyer MC**, **Seveno M**, **Pagny-Salehabadi S**, **Bardor M**, **Durambur G**, **Lerouge P**, **Vezina L**, **Gomord V**; Down-regulated expression of plant-specific glycoepitopes in alfalfa. Plant Biotechnol J. **2008**; 6(7):702-21.

Smith PL, **Gersten KM**, **Petryniak B**, **Kelly RJ**, **Rogers C**, **Natsuka Y**, **Alford JA 3rd**, **Scheidegger EP**, **Natsuka S**, **Lowe JB**; Expression of the α(1,3)fucosyltransferase Fuc-TVII in lymphoid aggregate high endothelial venules correlates with expression of L-selectin ligands. J Biol Chem. **1996**; 271(14):8250-9.

Springer TA; Traffic signals for lymphocyte recirculation and leukocyte emigration: the multistep paradigm. Cell. **1994**; 76(2):301-14.

Stacke C, Ziegelmüller P, Hahn U; Comparison of expression systems for human fucosyltransferase IX. Eur J Cell Biol. **2009**

Staudacher E; α1,3-Fucosyltransferases. Trends Glycosci Glycotechn. **1996**; 8:391-408.

Sun HY, **Lin SW**, **Ko TP**, **Pan JF**, **Liu CL**, **Lin CN**, **Wang AH**, **Lin CH**; Structure and mechanism of Helicobacter pylori fucosyltransferase. A basis for lipopolysaccharide variation and inhibitor design. J Biol Chem. **2007**; 282(13):9973-82.

Takamiya K, **Yamamoto A**, **Furukawa K**, **Yamashiro S**, **Shin M**, **Okada M**, **Fukumoto S**, **Haraguchi M**, **Takeda N**, **Fujimura K**, **Sakae M**, **Kishikawa M**, **Shiku H**, **Furukawa K**, **Aizawa S**; Mice with disrupted GM2/GD2 synthase gene lack complex gangliosides but exhibit only subtle defects in their nervous system. Proc Natl Acad Sci USA. **1996**; 93(20):10662-7.

Takamiya K, **Yamamoto A**, **Furukawa K**, **Zhao J**, **Fukumoto S**, **Yamashiro S**, **Okada M**, **Haraguchi M**, **Shin M**, **Kishikawa M**, **Shiku H**, **Aizawa S**, **Furukawa K**; Complex gangliosides are essential in spermatogenesis of mice: possible roles in the transport of testosterone. Proc Natl Acad Sci USA. **1998**; 95(21):12147-52.

Tamura K, **Dudley J**, **Nei M**, **Kumar S**; MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. **2007**; 24(8):1596-9.

Tanaka S, **Yazawa S**, **Noguchi K**, **Nishimura T**, **Miyanaga K**, **Kochibe N**, **Poland DC**, **Van Dijk W**, **Matta KL**; Molecular analysis of plasma α1,3-fucosyltransferase deficiency and development of the methods for its genotyping. Exp Clin Immunogenet. **2001**; 18(1):1-12.

Thanka Christlet TH, **Veluraja K**; Database analysis of O-glycosylation sites in proteins. Biophys J. **2001**; 80(2):952-60.

Thompson JD, **Gibson TJ**, **Higgins DG**; Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics. **2002**; Chapter 2:Unit 2.3.

Tian E, **Ten Hagen KG**; Recent insights into the biological roles of mucin-type O-glycosylation. Glycoconj J. **2009**; 26(3):325-34.

Tsuchiya N, **Endo T**, **Matsuta K**, **Yoshinoya S**, **Aikawa T**, **Kosuge E**, **Takeuchi F**, **Miyamoto T**, **Kobata A**; Effects of galactose depletion from oligosaccharide chains on immunological activities of human IgG. J Rheumatol. **1989**; 16(3):285-90.

Tusnády GE, **Simon I**; The HMMTOP transmembrane topology prediction server. Bioinformatics. **2001**; 17(9):849-50.

Unligil UM, **Zhou S**, **Yuwaraj S**, **Sarkar M**, **Schachter H**, **Rini JM**; X-ray crystal structure of rabbit N-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily. EMBO J. **2000**; 19(20):5269-80.

Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J; Gold Spring Harbor Laboratory Press. Essentials of Glycobiology. **1999**.

Wandall HH, **Hassan H**, **Mirgorodskaya E**, **Kristensen AK**, **Roepstorff P**, **Bennett EP**, **Nielsen PA**, **Hollingsworth MA**, **Burchell J**, **Taylor-Papadimitriou J**, **Clausen H**; Substrate specificities of three members of the human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3. J Biol Chem. **1997**; 272(38):23503-14.

Wang J, **Wang W**, **Kollman PA**, **Case DA**; Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model. **2006**; 25(2):247-60.

Wang Y, **Shao L**, **Shi S**, **Harris RJ**, **Spellman MW**, **Stanley P**, **Haltiwanger RS**; Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase. J Biol Chem. **2001**; 276(43):40338-45.

Webster JC, **Klingman JD**; Incorporation of radiolabelled sugars into synaptic junctional macromolecules from chick brain. Brain Res Bull. **1980**; 5(1):31-4.

Wells L, **Vosseller K**, **Hart GW**; Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. Science. **2001**; 291(5512):2376-8.

Wiese TJ, **Dunlap JA**, **Yorek MA**; Effect of L-fucose and D-glucose concentration on L-fucoprotein metabolism in human Hep G2 cells and changes in fucosyltransferase and α-L-fucosidase activity in liver of diabetic rats. Biochim Biophys Acta. **1997**; 1335(1-2):61-72.

Wiggins CA, **Munro S**; Activity of the yeast MNN1 α1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. Proc Natl Acad Sci USA. **1998**; 95(14):7945-50.

Wilson IBH, **Harthill JE**, **Mullin NP**, **Ashford DA**, **Altmann F**; Core α1,3-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. **1998**; 8(7):651-61.

Wilson IBH, **Rendić D**, **Freilinger A**, **Dumić J**, **Altmann F**, **Mucha J**, **Müller S**, **Hauser MT**; Cloning and expression of cDNAs encoding α1,3-fucosyltransferase homologues from Arabidopsis thaliana. Biochim Biophys Acta. **2001**; 1527(1-2):88-96.

Wilson IBH; Glycosylation of proteins in plants and invertebrates. Curr Opin Struct Biol. **2002**; 12(5):569-77.

Woodcock S, **Mornon JP**, **Henrissat B**; Detection of secondary structure elements in proteins by hydrophobic cluster analysis. Protein Eng. **1992**; 5(7):629-35.

Wrabl JO, **Grishin NV**; Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. J Mol Biol. **2001**; 314(3):365-74.

Xu Z, **Vo L**, **Macher BA**; Structure-function analysis of human α1,3-fucosyltransferase. Amino acids involved in acceptor substrate specificity. J Biol Chem. **1996**; 271(15):8818-23.

Yazawa S, **Madiyalakan R**, **Izawa H**, **Asao T**, **Furukawa K**, **Matta KL**; Cancer-associated elevation of α(1→3)-L-fucosyltransferase activity in human serum. Cancer. **1988**; 62(3):516-20.

Yazawa S, **Asao T**, **Nagamachi Y**, **Abbas SA**, **Matta KL**; Tumor-related elevation of serum α(1→3)-L-fucosyltransferase activity in gastric cancer. J Cancer Res Clin Oncol. **1989**; 115(5):451-5.

Zachara NE, **Hart GW**; Cell signaling, the essential role of O-GlcNAc! Biochim Biophys Acta. **2006**; 1761(5-6):599-617.

7. Abstract

This work brings an insight into conserved motifs and sequence similarities of α1,3/4-fucosyltransferases with an emphasis on plant core α1,3-fucosyltransferases. Core α1,3-fucosylation is a conserved feature of plant and invertebrate *N*-linked oligosaccharides with no yet known biological function. Core α1,3-fucosyltransferase activity is connected to pollen, food and insect allergies in many patients. Core α1,3-fucose is recognised by the immune system as foreign element, since mammals do not produce this epitope.

I present the results of biochemical characterisation (such as divalent metal ion dependency and K_m for the donor substrate), protein truncation (deletion of the proposed most C-terminal subdomain unique to plant core α1,3-fucosyltransferases), protein fusions (of immuno-detection epitopes and His-tag), and site-directed mutagenesis of *Arabidopsis thaliana* core α1,3-fucosyltransferase A (FucTA). Mutagenesis was carried out by alanine replacement of potentially important residues based on the results of bioinformatics analyses and homology modelling (which is also subject of this work). Alanine screening was performed either to investigate the importance of some residues of the donor substrate binding pocket of *A. thaliana* FucTA or its putative *N*-glycosylation sites. The latter study was complemented also with mass spectral fingerprinting.

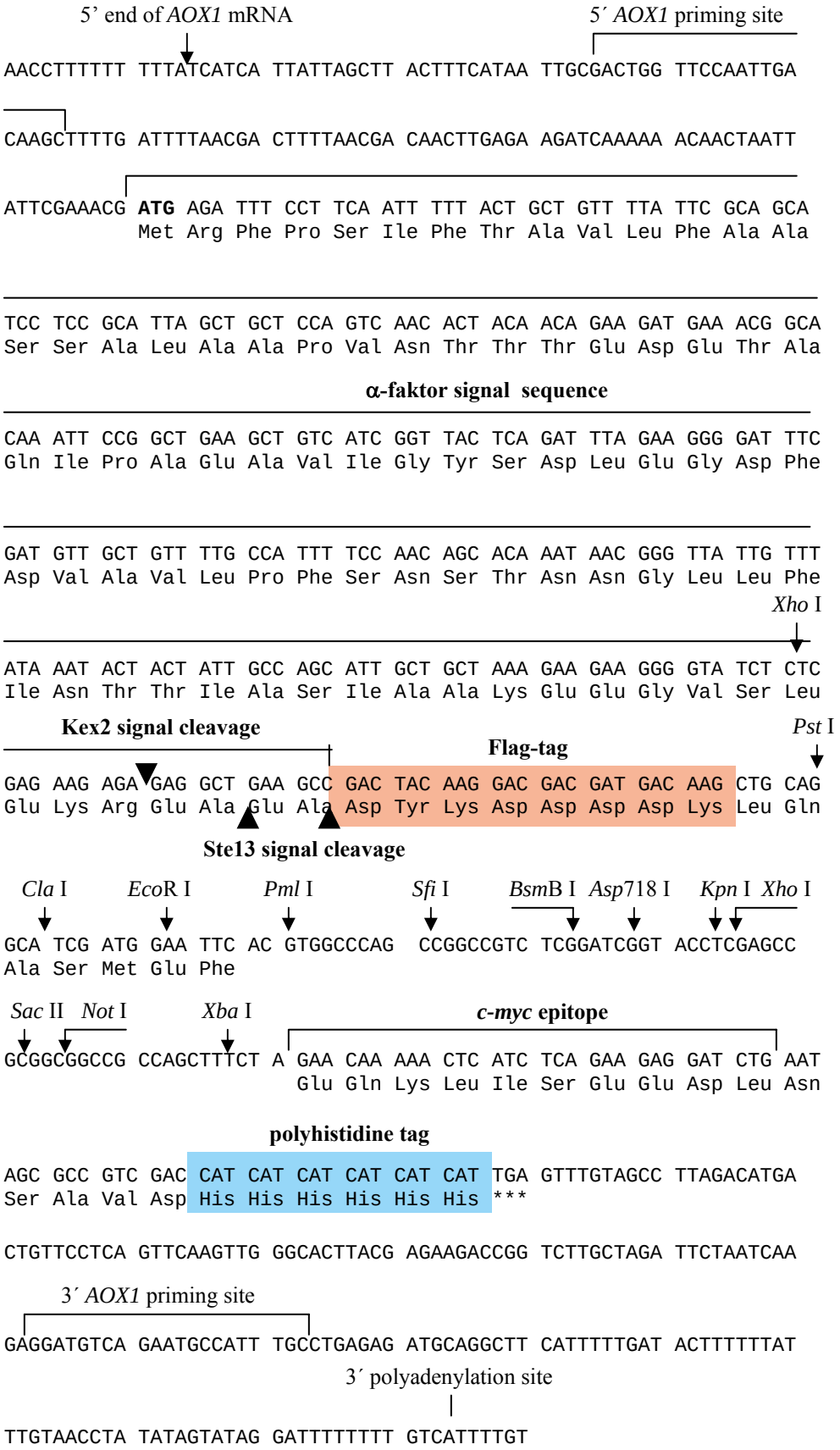
Key words: core α1,3-fucosyltransferases, *N*-glycans, fucosylation, structure-function studies, plant, nematode, GT-B fold, donor substrate binding, homology modelling

Ce travail cherche à apporter un éclairage sur les relations séquence-structure-fonction des α1,3/4-fucosyltransférases (famille GT10 des glycosyltransférases), avec un accent particulier sur les core α1,3-fucosyltransférases des plantes. La fucosylation de type Core α1,3 est une caractéristique des oligosaccharides *N*-liés des plantes et invertébrés, avec une fonction biologique qui n'est pas encore élucidée. L'activité Core α1,3-fucosyltransférase est responsable d'allergies alimentaires, au pollen, et aux insectes chez l'homme. En effet, le Core α1,3-fucose est reconnu par le système immunitaire comme un élément étranger car les mammifères ne produisent pas cet épitope.

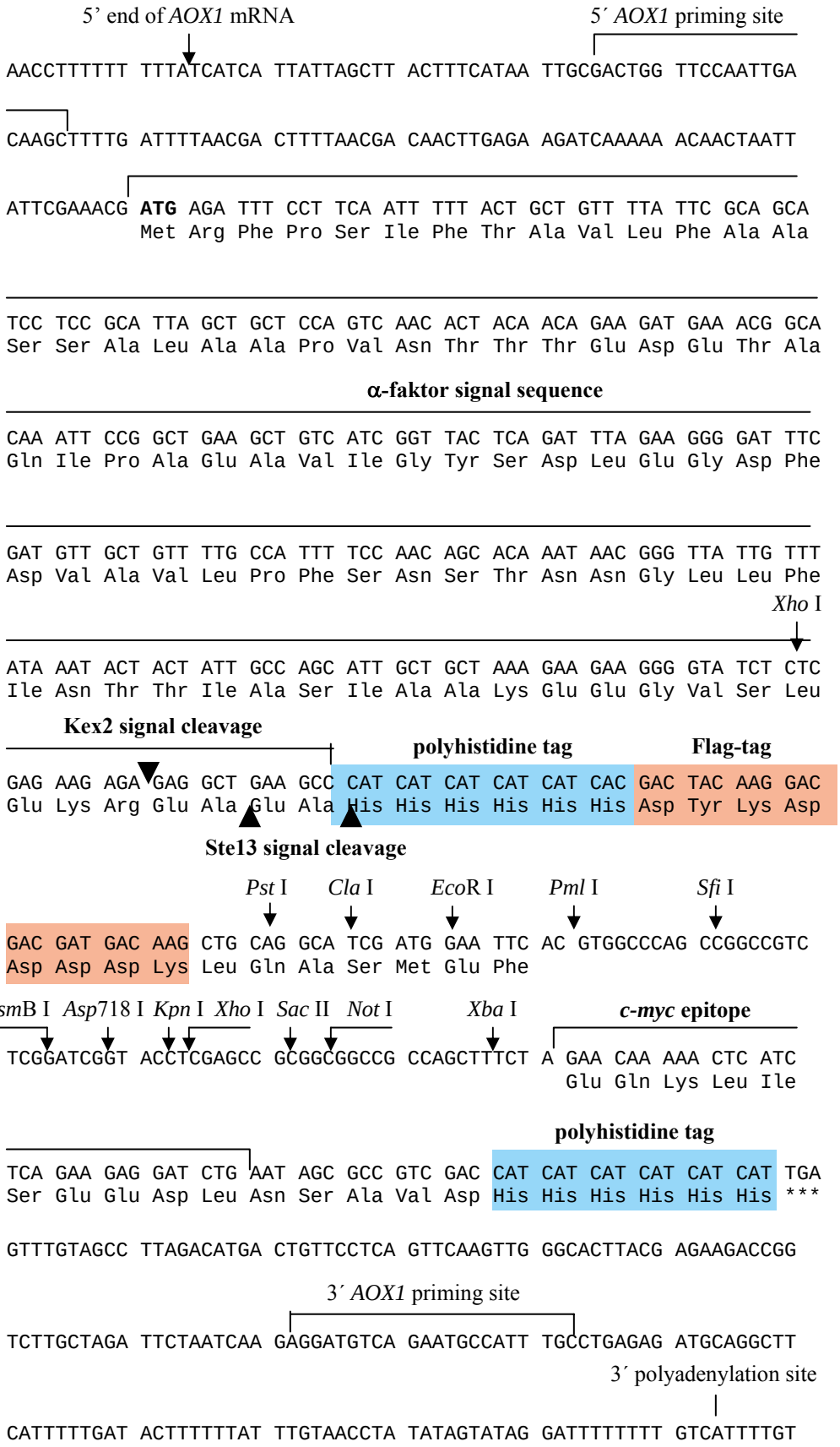
Dans le cadre de ce travail sont présentés des résultats de caractérisation biochimique (comme l'influence de cations divalents sur l'activité et du K_m de substrat donneur), des expériences de troncation des différents domaines (comme la suppression du sous-domaine C-terminal spécifique aux core α1,3-fucosyltransférases des plantes), de fusion (à des épitopes de type His-tag ou Flag) et de mutagenèse dirigée, en utilisant comme protéine modèle, la core α1,3-fucosyltransferase A (FucTA) d'*Arabidopsis thaliana* qui a été exprimée sous forme recombinante chez *Pichia pastoris*. Ces expériences ont été dictées sur la base de nos résultats d'analyses bioinformatiques des séquences de α1,3/4-fucosyltransférases et de la modélisation par homologie du domaine de liaison au nucléotide-sucre de l'enzyme FucTA. La mutagenèse des résidus clé identifiés par cette approche a permis de confirmer l'importance de certains acides aminés dans le mécanisme catalytique. Enfin la protéine FucTA étant elle-même glycosylée quand elle est produite chez *P. pastoris*, nous avons étudié l'impact de cette glycosylation sur la production et l'activité de la protéine, par des expériences de mutagenèse, de Western blotting et de spectrométrie de masse.

Appendix

pPICZαFlag C3



pPICZα HisFlag



Filename: Dissertation final correct text v_2.doc
Directory: C:\Documents and Settings\Soňa Garajová\Desktop
Template: C:\Documents and Settings\Soňa Garajová\Application Data\Microsoft\Templates\Normal.dot
Title: Theory
Subject:
Author: Matilda
Keywords:
Comments:
Creation Date: 1/18/2010 1:52:00 PM
Change Number: 2
Last Saved On: 1/18/2010 1:52:00 PM
Last Saved By: Soňa Garajová
Total Editing Time: 1 Minute
Last Printed On: 1/18/2010 1:56:00 PM
As of Last Complete Printing
Number of Pages: 129
Number of Words: 58 638 (approx.)
Number of Characters: 334 243 (approx.)